Multiple disease detection from Chest X-Rays and storing the records in Blockchain using Deep Learning Techniques

Prof.Vijaykumar Bhanuse, Shreya Singh, Siddhesh Shinkar, Harshal Dharap, Aneesh Poduval

Department of Instrumentation Engineering

Vishwakarma Institute of Technology Pune, India.

Abstract— Lung disease refers to several types of diseases or disorders that prevent the lungs from functioning properly. Lung disease can affect respiratory function, or the ability to breathe, and pulmonary function, which is how well lungs work. There are many different lung diseases, some of which are caused by bacterial, viral, or fungal infections. Other lung diseases are associated with environmental factors, including asthma, mesothelioma, and lung cancer. Chronic lower respiratory diseases is a set of conditions that includes chronic obstructive pulmonary disease (COPD), emphysema, and chronic bronchitis. Together, chronic lower respiratory diseases are a leading cause of death in the United States. Respiratory diseases such as asthma and COPD involve a narrowing or blockage of airways that reduce air flow. In other lung conditions– such as pulmonary fibrosis, a lung tissue scarring that can be caused by different factors, and pneumonia, a bacterial or viral infection in which air sacs fill with fluid–the lungs have reduced ability to hold air. Through this research project, we seek to contribute to the medical field by implementing deep learning technology along with machine learning algorithms into a system which can detect multiple types of chest/ lung diseases from X-rays reports quickly and store the resulting records in a Blockchain system.

Keywords : deep learning, blockchain, lung disease, medical images.

I. INTRODUCTION

Chest X-ray exams are one of the most frequent and cost-effective medical imaging examinations available. However, clinical diagnosis of a chest X-ray can be challenging and sometimes more difficult than diagnosis via chest CT imaging. The lack of large publicly available datasets with annotations means it is still very difficult, if not impossible, to achieve clinically relevant computer-aided detection and diagnosis (CAD) in real world medical sites with chest X-rays. One major hurdle in creating large X-ray image datasets is the lack resources for labeling so many images [1]. Technology today is progressing in ways that make detection of such conditions faster, less invasive and more accessible to ordinary citizens. This is just one of the many examples of the amazing performance in the hea systems brought about by recent advancements in technology, which raises the bar for traditional computational systems as well.

"Deep learning" is a subset of machine learning which primarily focuses on artificial neural networks, which aim to replicate the way the human brain processes data via "representation learning". Advancements in computing technology on both hardware and software fronts have led to a boom in popularity of deep learning among enthusiasts and professionals alike.

In recent times, deep learning methods such as Convolutional neural networks have made enormous progress in the medical field, most notably in cases involving medical imaging. Given enough computational power, they have been shown to even outperform human experts in their relevant areas of expertise.

Through this research project, we have implemented a deep learning convolutional neural network system which, after suitable feature engineering and hyperparameter tuning, can detect multiple chest diseases after analysing with an accuracy of 75% storing the results in a blockchain system after generating a smart contract.

II. LITERATURE SURVEY

In the present-day scenario, alot of work is being performed in the healthcare field to relieve people of their difficulties. We came across a large number of papers in this process which aid to detect the presence of diseases in X-rays especially in terms of tumour classification. Several papers used a traditional machine learning approach [2, 3, 4, 5, 6], while others utilized a more complex deep learning approach [7, 8, 9]. Some have even applied additional signal processing techniques to enhance model performance. Deep learning methods predictably outperform traditional algorithms in most cases, with the obvious trade-off of computational complexity, especially in the case of convolutional neural networks. From the literature, a majority of the works surveyed used features that are automatically extractedfrom CNN. CNN can automatically learn and extract features, discarding the need for manual feature generation.

Detection of multiple disease classification and storing the resulting records using a Blockchain is unique to our paper.

III. DATASET

This dataset was taken from the Kaggle dataset, consisting of NIH chest X-rays [10]. The dataset consisting of multiple chest X-rays of various different kinds of patients. The accurate count of total number of chest X-rays that were dealt during the making of the project is 1,12,000 and number of patients exceeded 30000 in number.

Further Exploratory Data Analysis on given data showed that 56 % of patients in the dataset were male while the remaining 44 % were female. Each patient had one ray X-ray that was denoted by a unique patient identity. The finding labels that were detected by the Convolutional Neural Network was divided into 15 classes namely :

- Atelectasis
- Consolidation
- Infiltration
- Pneumothorax
- Edema
- Emphysema
- Fibrosis
- Effusion
- Pneumonia
- Pleural_thickening
- Cardiomegaly
- Nodule Mass
- Hernia

One class has been named as 'No Findings' in the dataset as no label was assigned since no detections were made in this case. On further exploratory data analysis it was found that just under 50 percent of the Chest X-Rays reported no findings in the dataset

The dataframe pertaining to the patient information consisted of several import columns such as Image Index, which was unique to each chest X-ray. Finding Labels: Disease type (Class label), Follow-up : whether of not the patient has taken a follow up, Patient ID, Patient Age , Patient Gender, View Position: X-ray orientation, Original Image Width Original Image Height, Original Image Pixel Spacing x Original Image Pixel Spacing y.

Data limitations

1. The image labels are NLP extracted so there could be some erroneous labels but the NLP labeling accuracy is estimated to be >90%.
2. Very limited numbers of disease region bounding boxes.
3. Chest x-ray radiology reports are not anticipated to be publicly shared, which makes dataset acquisition as well as image annotation some of the challenges that are faced in the medical field.

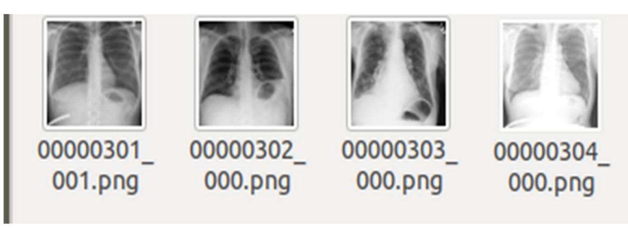


Figure 1: Images in the dataset

IV. METHODOLOGY

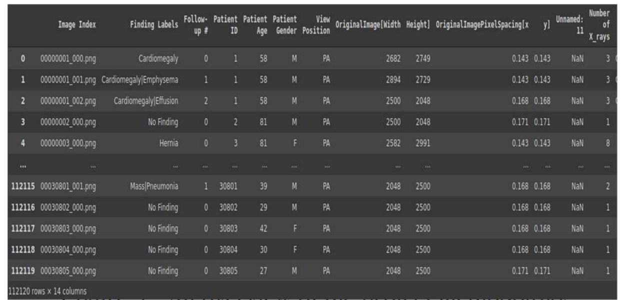


	Image Index	Finding Labels	Follow-up #	Patient ID	Patient Age	Patient Gender	View Position	OriginalImage[Width	Height]	OriginalImagePixelSpacing[x	y]	Unnamed: 11	Number of X_rays
0	00000001_000.png	Cardiomegaly	0	1	58	M	PA	2682	2749	0.143	0.143	NaN	3
1	00000001_001.png	Cardiomegaly|Emphysema	1	1	58	M	PA	2894	2729	0.143	0.143	NaN	3
2	00000001_002.png	Cardiomegaly|Effusion	2	1	58	M	PA	2500	2048	0.168	0.168	NaN	3
3	00000002_000.png	No Finding	0	2	81	M	PA	2500	2048	0.171	0.171	NaN	1
4	00000003_000.png	Hernia	0	3	81	F	PA	2582	2991	0.143	0.143	NaN	8
...	...	...	...	...	...	...	...	...	...	...	...	...	...
112115	00030801_001.png	Mass|Pneumonia	1	30801	39	M	PA	2048	2500	0.168	0.168	NaN	2
112116	00030802_000.png	No Finding	0	30802	29	M	PA	2048	2500	0.168	0.168	NaN	1
112117	00030803_000.png	No Finding	0	30803	42	F	PA	2048	2500	0.168	0.168	NaN	1
112118	00030804_000.png	No Finding	0	30804	30	F	PA	2048	2500	0.168	0.168	NaN	1
112119	00030805_000.png	No Finding	0	30805	27	M	PA	2048	2500	0.171	0.171	NaN	1

112120 rows × 14 columns

Illustration 2: Overview of the dataframe

After acquiring the dataset, we conduct exploratory data analysis and data pre-processing to make it suitable for model ingestion. We then apply multiple machine learning and deep learning models to establish a baseline performance. After evaluating the models, we tune the hyperparameters of the best performing one to improve its performance and then integrate it into a system to generate predictions from new data.

The final predictions are used by the doctor when a patient registers with him at the hospital and a smart contract is generated.

The model selection was done amongst three different kinds of models. These were inception networks, convolutional neural networks and ResNets. For the final outcome our CNN performed the best in terms of classifying the Chest X-rays with maximum accuracy and hence was the best performing model for the given dataset.

V. EXPLORATORY DATA ANALYSIS (EDA)

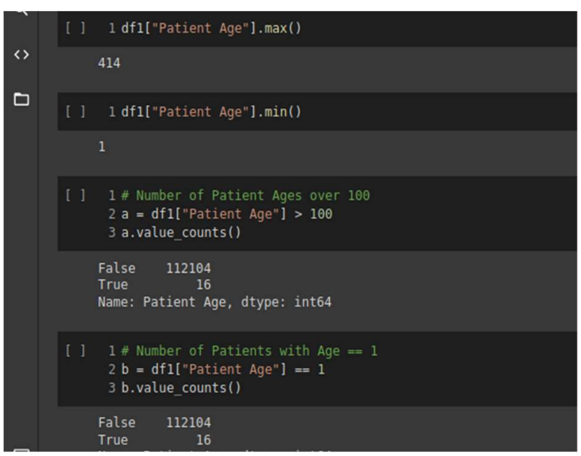
On carrying EDA alot of biases as well as outliers were detected, while carrying out the programming process. Alot of discrepencies were noticed in the Data itself. The maximum patient age according to the Data was found to be 414 and the minimum was 1. More than 16 patients in the Dataset recorded a patient age of 100 which was a clear outlier in terms of Data Visualization.

The first step here was dealing with bad age. Removing that data would result in information loss. To get a more accurate respresentation of the data we filled the outlier patient ages with the mean of the ages that were present in the column. We now had a more well balanced Data in terms of the patient age. The mean patient age was now found to be 48 years.

```
1 #Number of Males and Females
2 df1["Patient Gender"].value_counts(normalize = True)

M    0.56493
F    0.43507
Name: Patient Gender, dtype: float64

1 df1["Patient Gender"].value_counts().plot(kind = "barh")

<matplotlib.axes._subplots.AxesSubplot at 0x7f1dcc496550>
```


Figure 4: EDA of Patient Gender

In terms of Patient gender, there was found to be no bias. 56 percent of the patients were found to be male while the rest 44 percent were female. Therefore, a near even distribution was found in terms of gender.

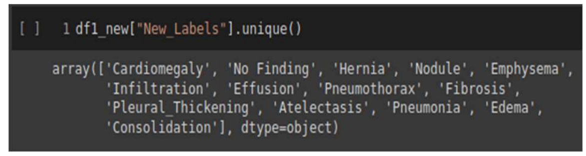


```
1 df1["Patient Age"].max()

414

1 df1["Patient Age"].min()

1

1 # Number of Patient Ages over 100
2 a = df1["Patient Age"] > 100
3 a.value_counts()

False    112104
True         16
Name: Patient Age, dtype: int64

1 # Number of Patients with Age == 1
2 b = df1["Patient Age"] == 1
3 b.value_counts()

False    112104
True         16
```

Figure 5: Bias in Patient Age

Furthermore, every finding classified the labels into more than one class. When the final number of labels had to be checked for, we saw that there were 836 unique labels instead of 15. The challenge was to compress these labels into 15 different classes. This was done by studying the distrubition of the labels for each unique patient ID. The diseases with the highest occurence in the finding labels was assigned to that particular patient X-ray.

Two more biases that were found where that 'No findings' comprised of almost 50 percent of the the diseases in the dataset. This was solved by augmenting the images that consisted of some labels that were assigned to them using the Data generator class of Keras. In this case we had more number of images added to the dataset and therefore larger information retrieval.

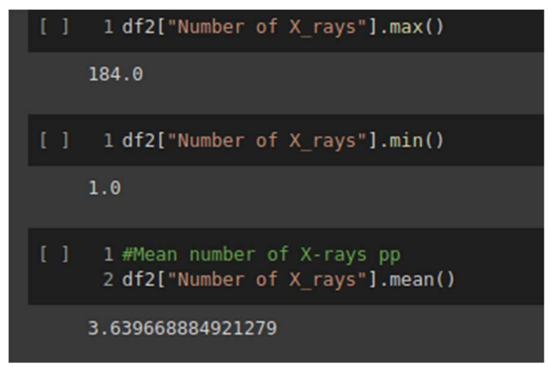


```
1 df1_new["New_Labels"].unique()

array(['Cardiomegaly', 'No Finding', 'Hernia', 'Nodule', 'Emphysema',
       'Infiltration', 'Effusion', 'Pneumothorax', 'Fibrosis',
       'Pleural_Thickening', 'Atelectasis', 'Pneumonia', 'Edema',
       'Consolidation'], dtype=object)
```

Figure 6: Final patient labels

The final bias handling was done for the number of X-Rays that was performed for each patient. While the mean number of X-rays per patient was somewhere around 4 there was an evident skewed distribution to the left in terms of X-rays since the maximum number of X-rays for a unique patient ID was found to ne 184.

Again we normalized by removing the number of X-rays for patients who had taken a high number of X-rays since discarding them would be cost effective as well as avoid inducing unnecessary bias into the model.



```
1 df2["Number of X_rays"].max()

184.0

1 df2["Number of X_rays"].min()

1.0

1 #Mean number of X-rays pp
2 df2["Number of X_rays"].mean()

3.639668884921279
```

Figure 7 : Bias in no of X-Rays

VI. MODEL ARCHITECTURES

CONVOLUTIONAL NEURAL NETWORK

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

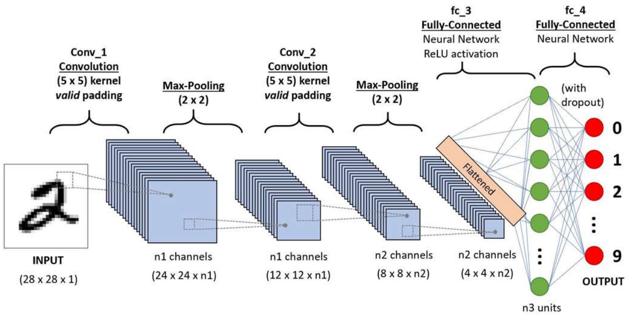


Figure 8 : Architecture of the CNN

INCEPTION NETWORK

Inception Layer) is a combination of all those layers (namely, 1×1 Convolutional layer, 3×3 Convolutional layer, 5×5 Convolutional layer) with their output filter banks concatenated into a single output vector forming the input of the next stage.

A layer in our deep learning model has learned to focus on individual parts of a face. The next layer of the network would probably focus on the overall face in the image to identify the different objects present there. Now to actually do this, the layer should have the appropriate filter sizes to detect different objects.

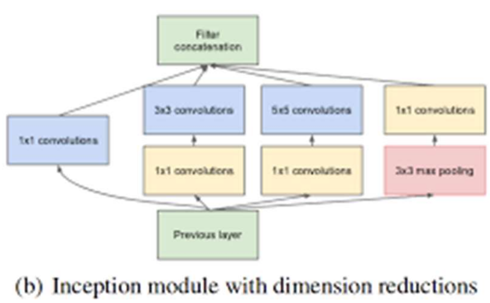


Figure 9: Architecture of Inception Network

This is where the inception layer comes to the fore. It allows the internal layers to pick and choose which filter size will be relevant to learn the required information. So even if the size of the face in the image is different (as seen in the images below), the layer works accordingly to recognize the face. For the first image, it would probably take a higher filter size, while it'll take a lower one for the second image.

Figure 10: Inception Nets

RESIDUAL NETWORK

ResNet, short for Residual Network is a specific type of neural network that was introduced in 2015 by Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun in their paper "Deep Residual Learning for Image Recognition".

Mostly in order to solve a complex problem, we stack some additional layers in the Deep Neural Networks which results in improved accuracy and performance. The intuition behind adding more layers is that these layers progressively learn more complex features. For example, in case of recognising images, the first layer may learn to detect edges, the second layer may learn to identify textures and similarly the third layer can learn to detect objects and so on. But it has been found that there is a maximum threshold for depth with the traditional Convolutional neural network model.

Residual Block

This problem of training very deep networks has been alleviated with the introduction of ResNet or residual networks and these Resnets are made up from Residual Blocks.

The very first thing we notice to be different is that there is a direct connection which skips some layers(may vary in different models) in between. This connection is called 'skip connection' and is the core of residual blocks. Due to this skip connection, the output of the layer is not the same now. Without using this skip connection, the input 'x' gets multiplied by the weights of the layer followed by adding a bias term.

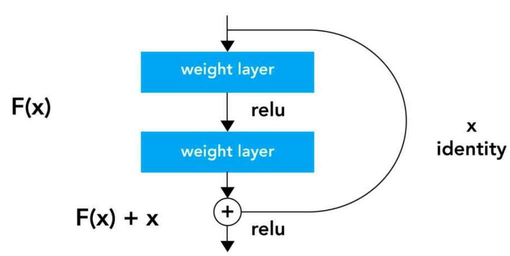


Figure 2. Residual learning: a building block.

Figure 11: Residual Block

VII. RESULTS

The Convolutional Neural Network outperfomed our Inception networks and Residual Network. The training data was split into 40 percent training and 60 percent testing to make sure it was generalizing well on the dataset.

The validation accuracy on the CNN was 75 percent after tuning the model upto 15 epochs and using the Adam Optimizer. The CNN architecture consisted of two convolutional layers with just over 43 million trainable parameters and no non trainable parameters.

In comparison, the Residual Network reached an accuracy of 62 percent and the Inception Network of 58 percent on the validation data. The validation loss for the Convolutional Neural Network Model was the least and kept on decreasing after each epoch. The minimum validation and training losses were noted after 15 epochs and were 1.36 and 1.76 respectively.

VIII. CONCLUSION

From the results, we can conclude that the Convolutional Neural Network model was the best choice since it showed the highest accuracy and was found to generalise the best in terms of the validation data. Though inception nets ans residual networks are more advanced algorithms they were found to be lacking in terms of accuracy and generalising well on the dataset.

These records were then stored in the Blockchain and integrated. Once the patient visits a hospital, he/she gets registered under a certain doctor and a smart contract is generated for that patient.

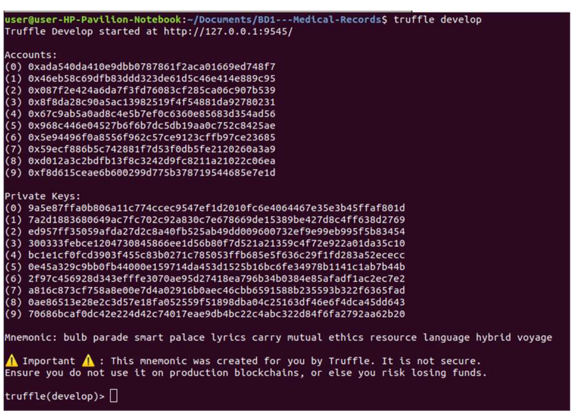


Figure 12: Blockchain integration

Figure 13: Results from Blockchain

IX. ACKNWOLEDGEMENT

- The authors would like to thank Prof. Vijaykumar Bhanuse for constant help and guidance.

X. REFERENCES

[1] https://www.niehs.nih.gov/health/topics/conditions/lung-disease/index.cfm.

[2] Sannino, Giovanna, and Giuseppe De Pietro. ”A deep learning approach for ECG-based heartbeat classification for arrhythmia detection.” Future Generation Computer Systems 86 (2018): 446-455.

[3] Ge, Dingfei, Narayanan Srinivasan, and Shankar M. Krishnan. ”Cardiac arrhythmia classification using autoregressive modeling.” Biomedical engineering online 1.1 (2002): 5.

[4] Batra, Anish, and Vibhu Jawa. ”Classification of arrhythmia using conjunction of machine learning algorithms and ECG diagnostic criteria.” Training Journal (1975).

[5] Soman, Thara, and Patrick O. Bobbie. ”Classification of arrhythmia using machine learning techniques.” WSEAS Transactions on computers 4.6 (2005): 548-552.

[6] Lyon, Aurore, et al. ”Distinct ECG phenotypes identified in hypertrophic cardiomyopathy using machine learning associated with arrhythmic risk markers.” Frontiers in physiology 9 (2018): 213.

[7] Gao, Dayong, et al. ”Arrhythmia identification from ECG signals with a neural network classifier based on a Bayesian framework.” Proceedings of the 24th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence. Citeseer, 2004.

[8] Oh, Shu Lih, et al. ”Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats.” Computers in biology and medicine 102 (2018): 278-287.

[9] Kurt Hornik (1991) “Approximation Capabilities of Multilayer Feedforward Networks”, Neural Networks, 4(2), 251–257. doi:10.1016/0893-6080(91)90009-T

[10] Dataset:- https://www.kaggle.com/nih-chest-xrays/data