Predictive Analysis of Lung Diseases using Machine Learning & Deep Learning Techniques

Group Id : GB7

Guide: Dr. Anuradha Thakare

Members :

Anant Sutar (B150334405)
Shravani Nimbolkar (B150334338)
Omkar Biranje (B150334340)
Subhradeep Mitra (B150334400)

Contents

INTRODUCTION

- Chest radiography is an important diagnostic tool for chest-related diseases. Medical imaging research is currently embracing the automatic detection techniques used in computer vision.
- The chest X-ray (CXR) is an easy, economical, and commonly adopted tool for diagnosing lung diseases.
- A radiologist interprets an X-ray as either normal or presenting a disease such as lung cancer, tuberculosis, or pneumonia.
- One of the most common chest diseases is pneumonia, a lung infection caused by viruses, bacteria, or fungi.
- Pneumonia becomes life-threatening in case of infants and adults.

PROBLEM STATEMENT

If an x-ray image of a subject’s lungs is fed into the machine learning model, then, the model should produce fast and accurate results about the patient suffering from the disease on which the model is trained on or not.

MOTIVATION

Lungs also play a major role in the functioning of the body. Hence it becomes very much necessary to ensure their proper functioning. Lung Diseases are presenting a huge challenge. The manual diagnostics of these kind of diseases, although possible, has some drawbacks, and thus we think it’s the need of the hour to implement automated systems, using the latest available techniques. The severity of chronic respiratory diseases can be understood by considering the very latest example of Covid-19. It has led to death of millions of people all over the world. Thus, we aim to address the problem by building automated software that could aid in the diagnostics of Lung Diseases.

OBJECTIVES

- ➔ Collection of Datasets
- ➔ Analyzing correlation
- ➔ Build robust models
- ➔ Analysis of different algorithms for Prediction of Lung Diseases
- ➔ Propose a novel architecture which is efficient & deliver good results

LITERATURE REVIEW

Title	Input with no.of subjects	Architecture	Problem	Output
(1)Synthetic Medical Images Using F&BGAN for Improved Lung Nodules Classification by VGG16	353 - 158 Benign , 183 Malignant	100-D input converted to 64x64 image by using 4 convolutional layers(Generator) O/P of Gen. turned to 100-D again by use of 4 Conv layers and followed by 2 FC layer(Disc.)	Limited availability and diversity of sample	Obtained accuracy of 94.24%
(2)Automated Pulmonary Node Classification in CT Images using DCNN trained by GAN.	60 cases out of which 27 were benign nodules and 33 cases were malignant.	100 uniformly drawn vectors as input to 4 fractionaly strided conv layers(Gen.) O/P of Gen given to Disc. consisting of 4 conv layers	Classification accuracy low to due less availability of data.	Obtained distinguishing acuracy of 66.7% for benign sample and 93.3% for malignant sample.

LITERATURE REVIEW

Title	Input with no.of subjects	Architecture	Problem	Output
(3)Efficient GAN-based Chest Radiographs (CXR) augmentation to diagnose coronavirus disease pneumonia	108948 X-ray images of 32,717 patients	Input and attribute are fed to the encoder which is followed by generator which also takes additional attributes and then output is given	Dataset was highly imbalanced which caused inaccurate and false predictions	Model achieved 89% accuracy in terms of GAN based synthetic data.
(4)LGAN: Lung Segmentation in CT Scans Using Generative Adversarial Network	dataset of 220 individual CT scans with two metrics: segmentation quality and shape similarity	Deep Deconvnet Network is trained to generate the lung mask while an Adversarial Network is trained to discriminate segmentation maps from the ground truth and the generator, generator learns lung segmentation of the input CT scans.	Network structure is not optimal Can extract higher level features but requires more data as well as more parameters & higher computation cost	Achieved mean accuracy of 92% of the proposed LGAN architecture

LITERATURE REVIEW

Title	Input with no.of subjects	Architecture	Problem	Output
(5)Segmentation of Lungs in Chest X-Ray Image Using Generative Adversarial Networks	3 Datasets. Shenzhen dataset: 200, 40. JSRT dataset: 200, 20, 20. Montgomery dataset: 110, 10, 18. Epochs: 350	Encoder - Decoder model, with 4 Discriminators.	Over-Segmentation and under segmentation.	Proposed model is able to achieve a dice-score of 0.9740, and IOU score of 0.943.

LITERATURE REVIEW

Title	Input with no.of subjects	Architecture	Problem	Output
(6)A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images	Guangzhou Women and Children's Medical Center dataset, Normal: 1346 Bacterial Pneumonia: 2538 Viral Pneumonia: 1345	Pre-trained models for feature selection: AlexNet DenseNet121 InceptionV3 ResNet18 GoogLeNet	Model can be improved by increasing dataset	Overall accuracy of 96.39%

LITERATURE REVIEW

Title	Input with no.of subjects	Architecture	Problem	Output
(7) CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning	ChestX-ray14 dataset consisting of 30,805 unique x-ray images	121-layer Dense Convolutional Network or DenseNet (CheXNet)	The diagnosis of different medical practitioner may differ as per their experiences	Overall accuracy of including all ailments was 84.14% (73.45 - 93.71)

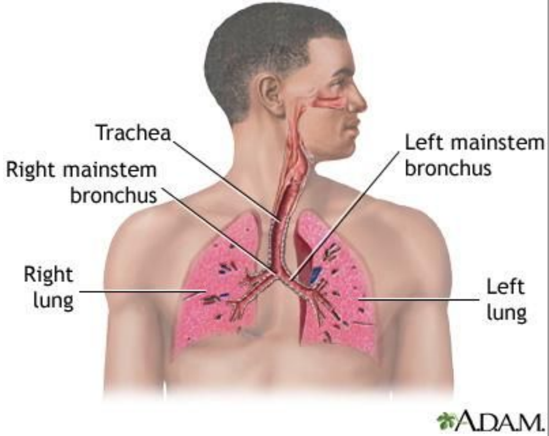
PLANNING: DATA ACQUISITION

- ➔ Lung disease: Any problem in the lungs that prevents the lungs from working properly.
- ➔ 3 main types of lung diseases:

  Airway diseases- Asthma, COPD and Bronchiectasis

  Lung tissue diseases - Pulmonary fibrosis and Sarcoidosis

  Lung circulation diseases – Pulmonary hypertension.



DATASET

- Pneumonia: Pneumonia is a lung infection that can range from mild to severe disease.
- It happens when an infection causes the air sacs in your lungs (your doctor will call them alveoli) to fill with fluid
- Dataset: Chest X ray images – https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

  https://www.kaggle.com/nih-chest-xrays/data

NORMAL

PNEUMONIC

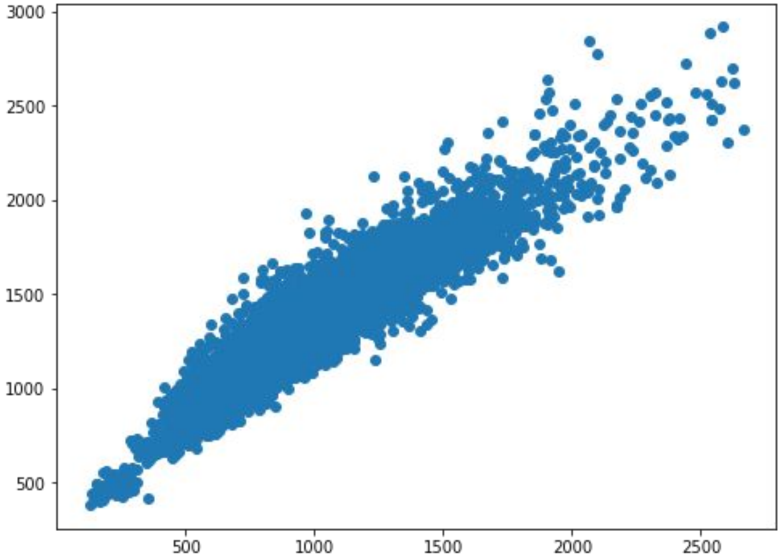
DATASET (Pneumonia)

Composition (total of 5856):

Normal: 1,583
Pneumonia: 4,273

Separated into:
4685 images for training
1171 images for validation

Image dimensions in dataset:



NIH dataset

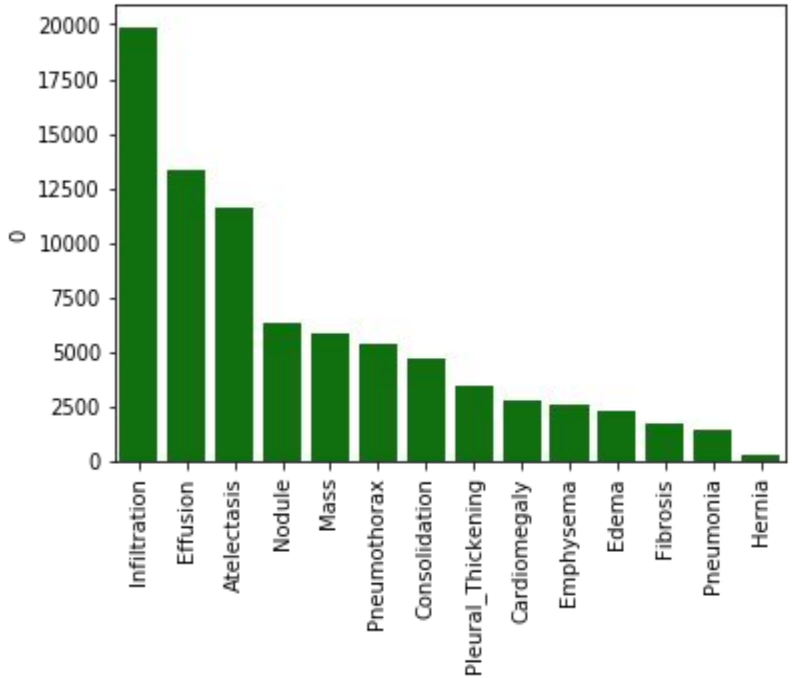
Dataset by : National Institutes of Health Chest X-Ray Dataset

Total samples : 112,120 X-ray images

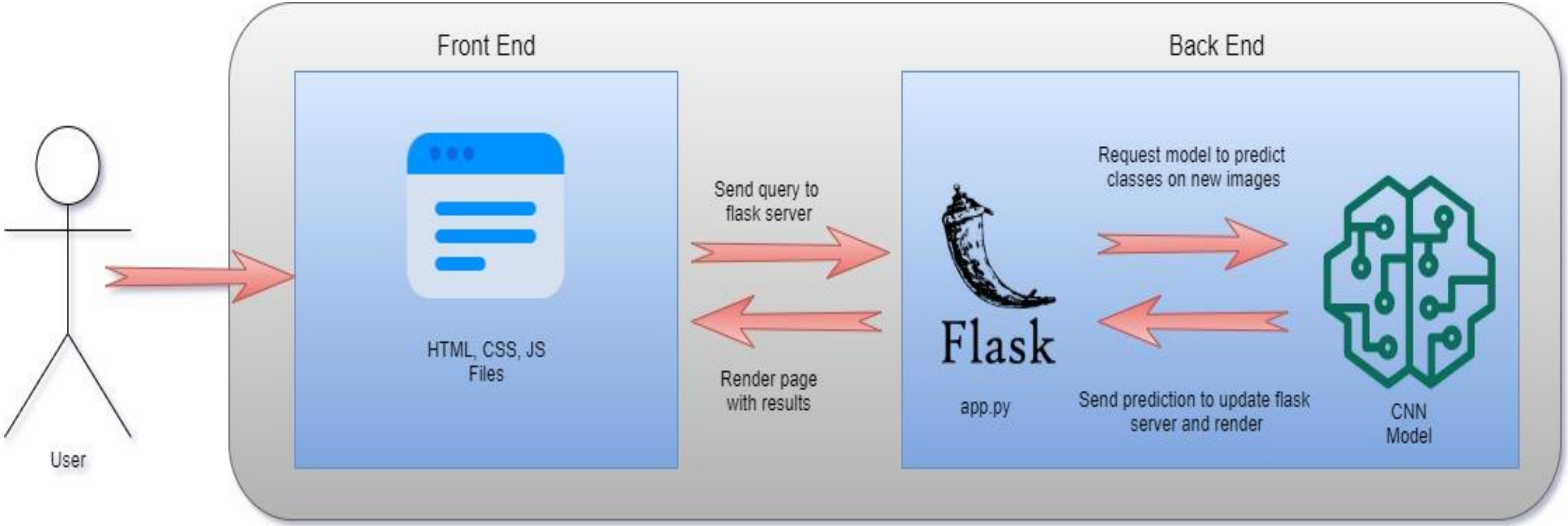
Size : 42 GB

Image size : 1024 X 1024

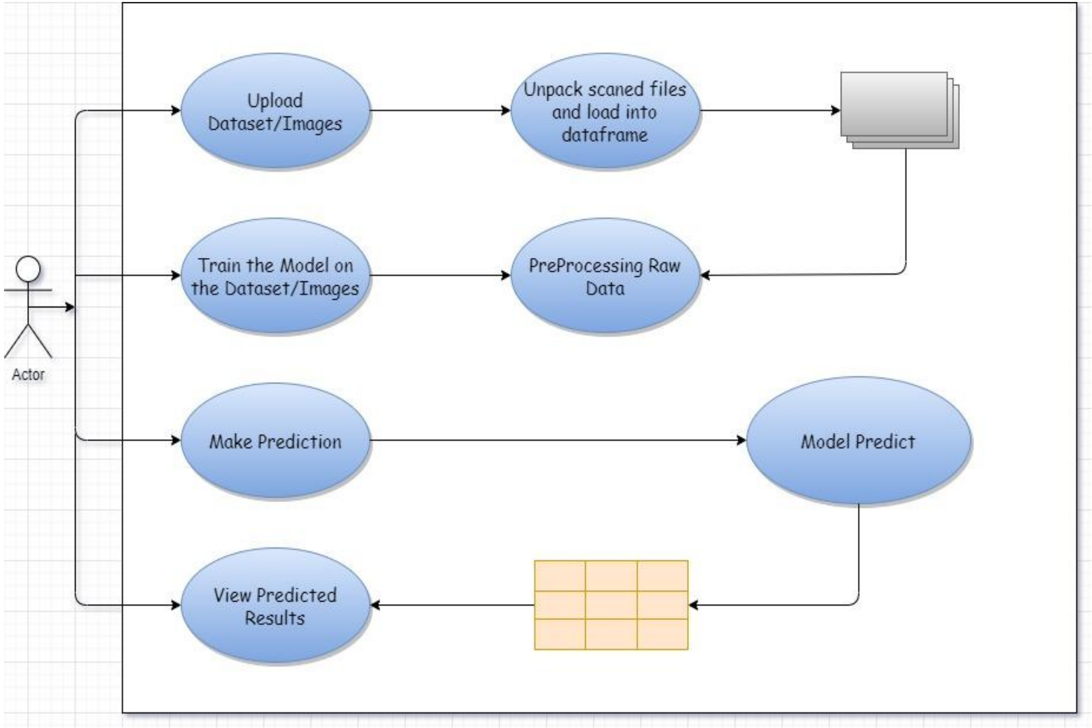
Total classes : 14 (lung disease) + 1(No findings)



SYSTEM ARCHITECTURE



USE CASE DIAGRAM



LIST OF ALGORITHMS

- → Naive bayes (Supervised)
- → Decision tree (Supervised)
- → Random forest (Supervised)
- → SVM (Supervised)
- → MLP (Supervised)
- → Ada Boost (Unsupervised)
- → Quadratic discriminant analysis QDA (Unsupervised)
- → VGG-16 (Transfer learning)
- → RESNET (Transfer learning)

Programming Tools and Platform

- Platform:
  - Tensorflow
  - Keras
  - OpenCV
- Model implementation:
  - Google Collab
  - Kaggle Notebook
- Data Storage:
  - File System

EXPERIMENTATION AND RESULTS

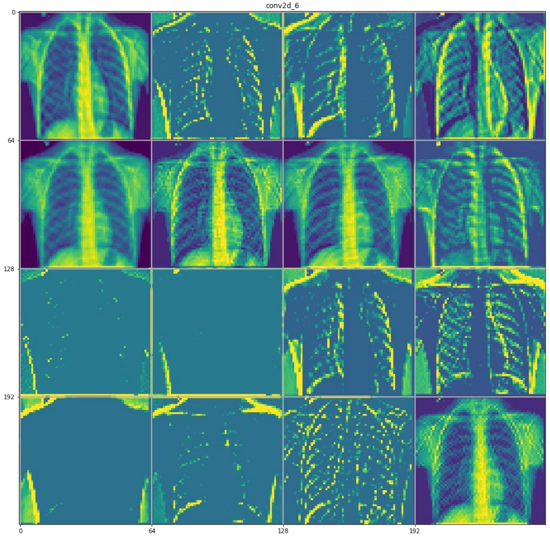
Candidate Models	Accuracy (Train)	Accuracy (Test)
Naive Bayes	85.18	72.92
Decision Tree	92.48	72.92
Random Forest	99.54	69.55
AdaBoost	97.18	71.96
MLP	99.54	75.96
SVM	98.24	76.44
Quadratic Discriminant Analysis	97.05	62.5
VGG 16	94.6	87.01
Resnet	85.2	82.13

EXPERIMENTATION AND RESULTS (contd.)

Candidate Models (CNN)	Accuracy (Train)	Accuracy (Test)
28 x 28	99.26	95.99
32 x 32	99.48	95.13
64 x 64	**99.46**	**96.13**
128 x 128	99.98	95.99
256 x 256	99.99	96.07

INTERMEDIATE ACTIVATIONS

Link: https://colab.research.google.com/drive/1i46UxNQcjt4mTeqGOxTbpFzf0EfxqwwF



ALGORITHMIC DETAILS & TECHNIQUE

Preprocessing:

1) Converting the images into grayscale format

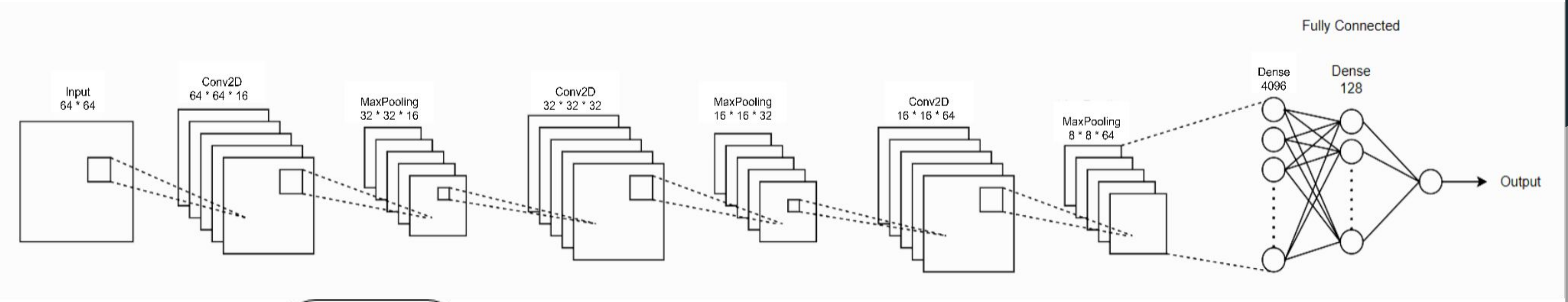
2) Resizing images into 64 * 64 size

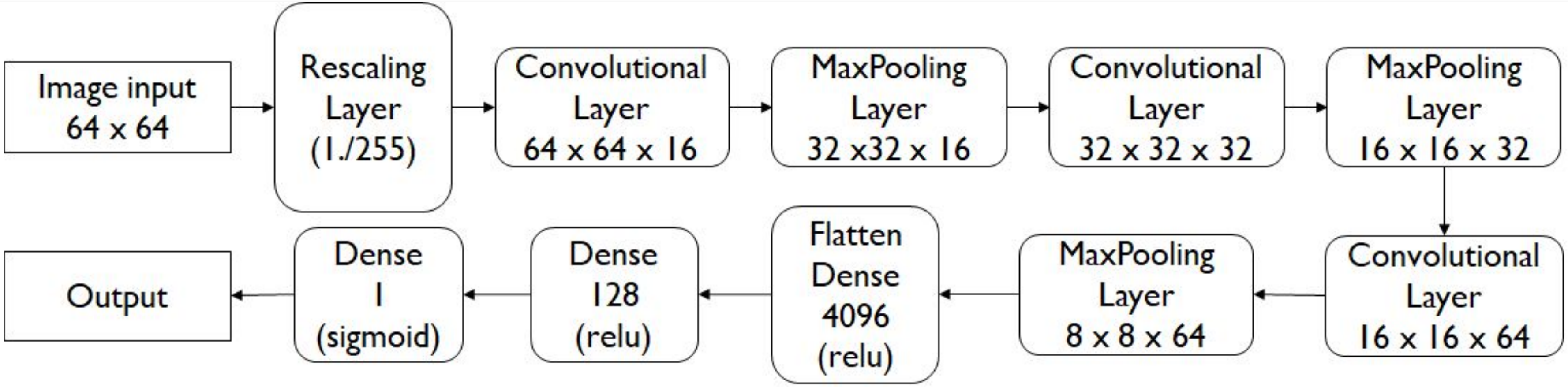
3) Applying Min-Max scaling technique to the values of each pixel

CNN MODEL:

- 3 2D Convolutional layers
- 3 Max pooling layers
- 1 flattening layer
- 1 Dense layer
- Output Dense layer

Model Architecture





RESULTS

Lung Disease	**Accuracy**
Atelectasis	83% https://bit.ly/3uqtUjO
Effusion	84% https://bit.ly/3xGU2ci
Pneumonia	96% https://bit.ly/3eT6jlD
No findings	80%
COVID(added later)	99%
Mass	80%
Nodule	78%
Pneumothorax	82%

APPLICATION

Web App Deployment

- ➔ The website introduces us to about 14 different diseases, which are a part of the NIH data. The user has to click on “Try Now” button in order to try out some images.
- ➔ The user will be on the upload image page where there will be a card, which will specify the disease on which the user wants prediction.
- ➔ On clicking on the specific disease he will have to upload the image of the Chest X-ray and click on Submit whereupon the user will have to wait for some time to get the results.

Technologies Used

- ➔ HTML, CSS , Bootstrap{ for designing purposes }
- ➔ Javascript { for validation and verification }
- ➔ PHP { for uploading images }
- ➔ Flask { for integrating ML model }
- ➔ Heroku { for deployment }

FLASK

- Flask is a Python web framework built with a small core and easy-to-extend philosophy.
- It is beginner friendly and comes with a boilerplate code and is also considered more pythonic than Django.
- It allows you to map routes (web addresses) to Python functions.
- It provides a few more pieces of functionality, such as sessions and template rendering.
- Since it is so lightweight, Flask can be better to get started quickly and understand fully what your application is doing.

HEROKU

- ➔ Heroku is a platform as a service (PaaS) that enables developers to build, run, and operate applications entirely in the cloud.
- ➔ Seamless GitHub integration means every pull request spins up a disposable Review App for testing, and any repo can be set up to auto-deploy with every GitHub push to a branch of your choosing.
- ➔ The Heroku network runs the customer's apps in virtual containers which execute on a reliable runtime environment. Heroku calls these containers "Dynos". These Dynos can run code written in Node, Ruby, PHP, Go, Scala, Python, Java, or Clojure
- ➔ Heroku lets the developer scale the app instantly just by either increasing the number of dynos or by changing the type of dyno the app runs in.

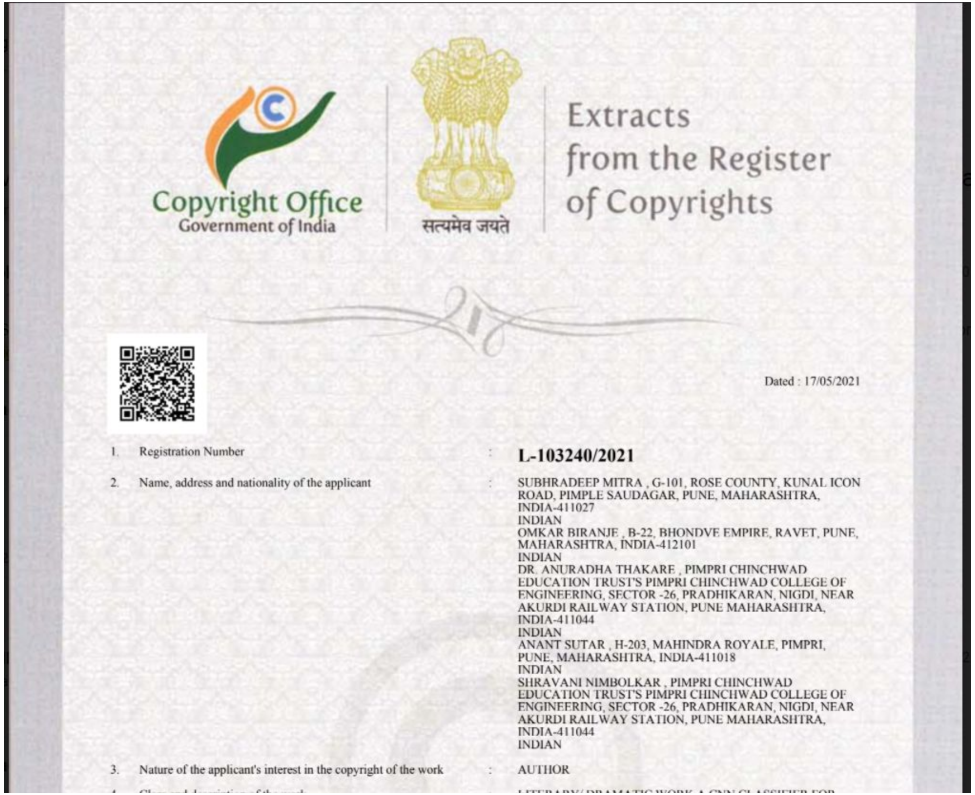
CONCLUSION

In this analytical study, we tested different approaches for the problem of pneumonia detection using AI & ML. We identified three different approaches namely machine learning, deep learning and transfer learning. CNN classifier among all the models gave the highest testing accuracy of 94.79% on Pneumonia disease. This is because the model is fine-tuned and it was able to generalize well on augmented data as well. Therefore, we planned to test the robustness of our CNN model on across range of diseases from NIH dataset. Our model was successfully able to classify the diseases and gave accuracy between 78-96% for various lung diseases like Atelectasis, lung nodule, Pneumothorax, COVID19 etc.

FUTURE SCOPE

Addition of Saliency maps to show the degree of infection and percentage of lung damage using Grad-Cam or Score-Cam for a better Diagnostics.

Copyright Status - Registered!



Copyright Office
Government of India

सत्यमेव जयते

Extracts from the Register of Copyrights

Dated : 17/05/2021

1.	Registration Number	:	L-103240/2021
2.	Name, address and nationality of the applicant	:	SUBHRADEEP MITRA , G-101, ROSE COUNTY, KUNAL ICON ROAD, PIMPLE SAUDAGAR, PUNE, MAHARASHTRA, INDIA-411027 INDIAN OMKAR BIRANJE , B-22, BHONDVE EMPIRE, RAVET, PUNE, MAHARASHTRA, INDIA-412101 INDIAN DR. ANURADHA THAKARE , PIMPRI CHINCHWAD EDUCATION TRUST'S PIMPRI CHINCHWAD COLLEGE OF ENGINEERING, SECTOR -26, PRADHIKARAN, NIGDI, NEAR AKURDI RAILWAY STATION, PUNE MAHARASHTRA, INDIA-411044 INDIAN ANANT SUTAR , H-203, MAHINDRA ROYALE, PIMPRI, PUNE, MAHARASHTRA, INDIA-411018 INDIAN SHRAVANI NIMBOLKAR , PIMPRI CHINCHWAD EDUCATION TRUST'S PIMPRI CHINCHWAD COLLEGE OF ENGINEERING, SECTOR -26, PRADHIKARAN, NIGDI, NEAR AKURDI RAILWAY STATION, PUNE MAHARASHTRA, INDIA-411044 INDIAN
3.	Nature of the applicant's interest in the copyright of the work	:	AUTHOR

DEMO

TITLE OF OUR PAPER

Detection of Lung Disease using Machine Learning and Deep Learning Techniques: An Analytical Study

Paper Publishing Details

Name of Journal : IGI Global as a Book Chapter

Paper Status : Under review

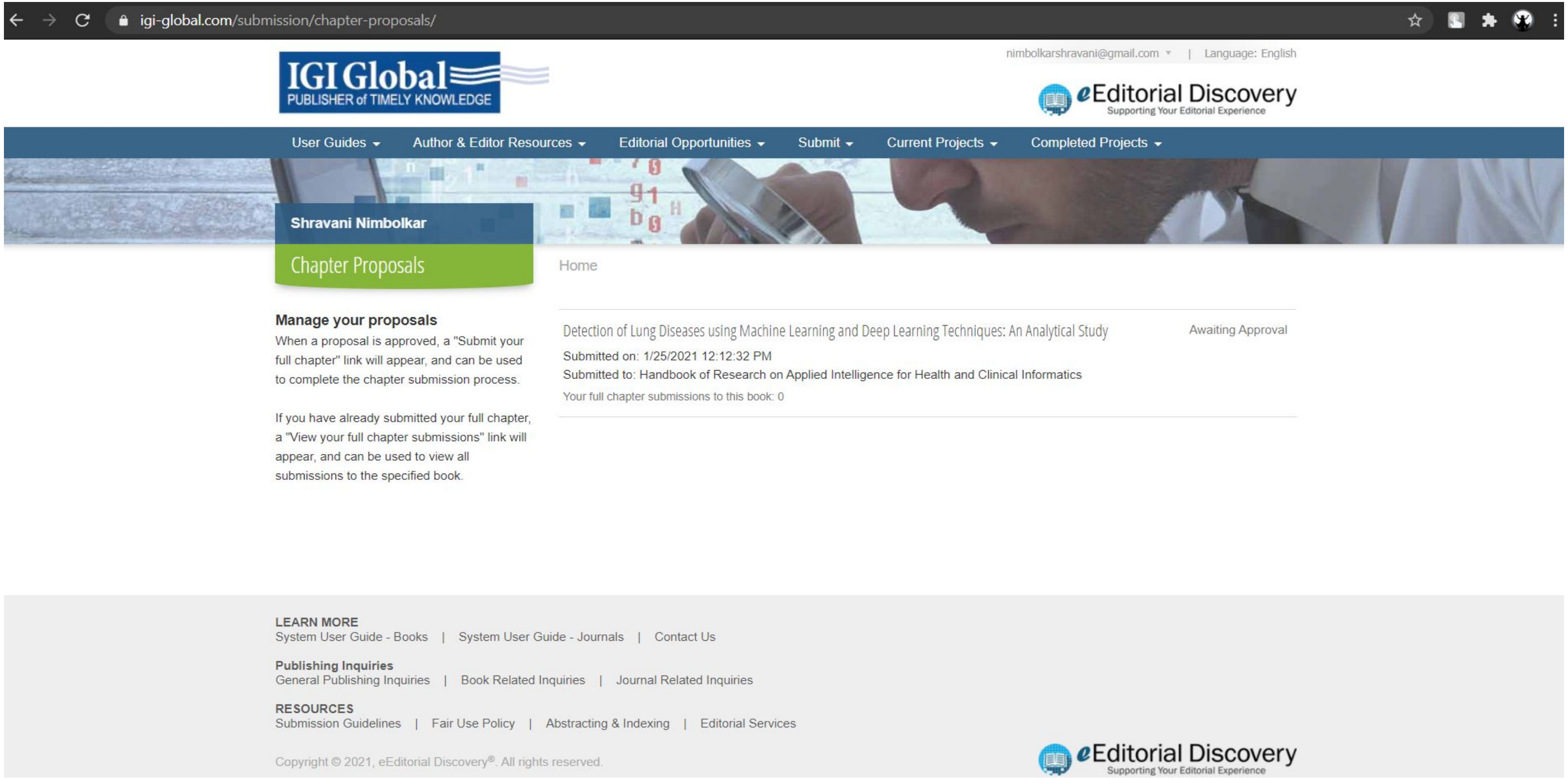
Submitted on : 25/01/2021

Reviewer comments : Improvements were suggested

Resubmitted on: 11/05/2021

Submitted to : Handbook of Research on Applied Intelligence for Health and Clinical Informatics

Proof of submission



igi-global.com/submission/chapter-proposals/

nimbolkarshravani@gmail.com | Language: English

IGI Global
PUBLISHER of TIMELY KNOWLEDGE

eEditorial Discovery
Supporting Your Editorial Experience

User Guides | Author & Editor Resources | Editorial Opportunities | Submit | Current Projects | Completed Projects

Shravani Nimbolkar

Chapter Proposals

Home

Manage your proposals

When a proposal is approved, a "Submit your full chapter" link will appear, and can be used to complete the chapter submission process.

If you have already submitted your full chapter, a "View your full chapter submissions" link will appear, and can be used to view all submissions to the specified book.

Detection of Lung Diseases using Machine Learning and Deep Learning Techniques: An Analytical Study

Awaiting Approval

Submitted on: 1/25/2021 12:12:32 PM

Submitted to: Handbook of Research on Applied Intelligence for Health and Clinical Informatics

Your full chapter submissions to this book: 0

LEARN MORE

System User Guide - Books | System User Guide - Journals | Contact Us

Publishing Inquiries

General Publishing Inquiries | Book Related Inquiries | Journal Related Inquiries

RESOURCES

Submission Guidelines | Fair Use Policy | Abstracting & Indexing | Editorial Services

eEditorial Discovery
Supporting Your Editorial Experience

Paper publication status

igi-global.com/submission/book-project-chapters/?projectid=c5b3fcd0-6e36-43b4-ad95-a3a107c1462d

(sw_anter@yahoo.com), Xiao-Zhi Gao (xiao.z.gao@gmail.com)

Release Date: November, 2021 | Copyright: © 2021 | Pages: 450

ISBN13: 9781799877097 | ISBN10: 1799877094 | EISBN13: 9781799877103

View Notifications | Visit Webpage

Progress report 2	May 30, 2021
Phase 3	
Revisions due from authors	Jun 27, 2021
Final acceptance/rejection notification due to authors	Jul 11, 2021
All final accepted materials due from authors	Jul 25, 2021
Final preface and table of contents	Aug 8, 2021
Final deadline	Sep 5, 2021

Select/deselect all | Total results: 1 | Sort by: Submitted (desc)

Filter Options

Chapter Submissions

Ch. N/A: Detection of Lung Diseases using Machine Learning and Deep Learning Techniques: An Analytical Study (020221-120617)

Chapter Information & Content

- Edit Chapter Title, Abstract, & Keywords
- Author Information (5)
- Documents (3)
- Figures (8)
- Overview

Assign & Manage Reviews

- Reviews Available to Author (0)

Phase History

- **Under Review** 3/22/2021 - *In Progress*
- Request Reviewers 3/22/2021 - 3/22/2021
- Remove Identifying Information 2/3/2021 - 3/22/2021
- Incomplete Submission 2/2/2021 - 2/3/2021

Show Notifications

LEARN MORE
System User Guide - Books | System User Guide - Journals | Contact Us

Publishing Inquiries
General Publishing Inquiries | Book Related Inquiries | Journal Related Inquiries

RESOURCES
Submission Guidelines | Fair Use Policy | Abstracting & Indexing | Editorial Services

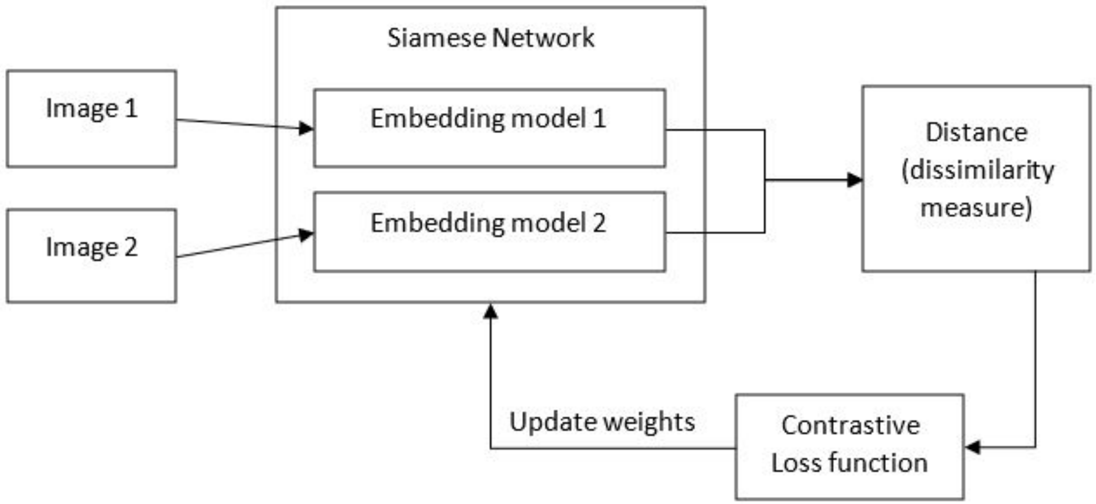
eEditorial Discovery
Supporting Your Editorial Experience

Link : https://www.igi-global.com/submission/book-project-chapters/?projectid=c5b3fcd0-6e36-43b4-ad95-a3a107c1462d

Algorithms we are working on

1) One shot learning
2) Few shot learning
3) Transfer learning
4) Hybrid model

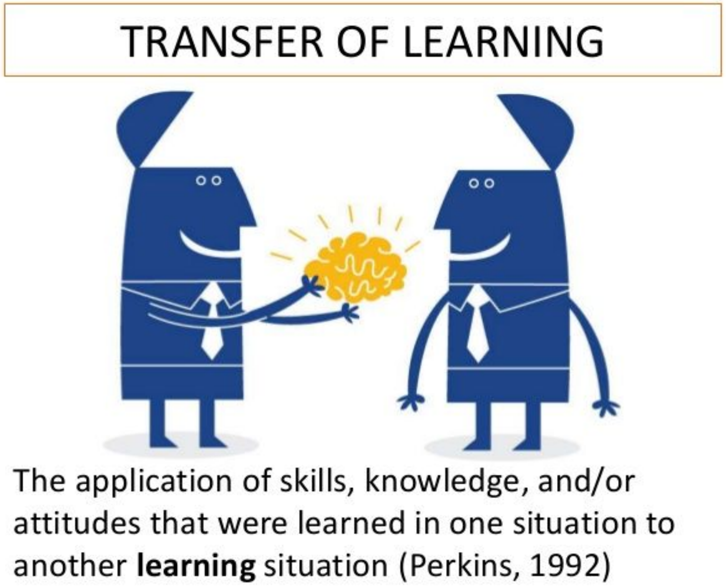
Structure of a simple Siamese Network



Transfer learning

- utilizing knowledge acquired for one task to solve related ones.

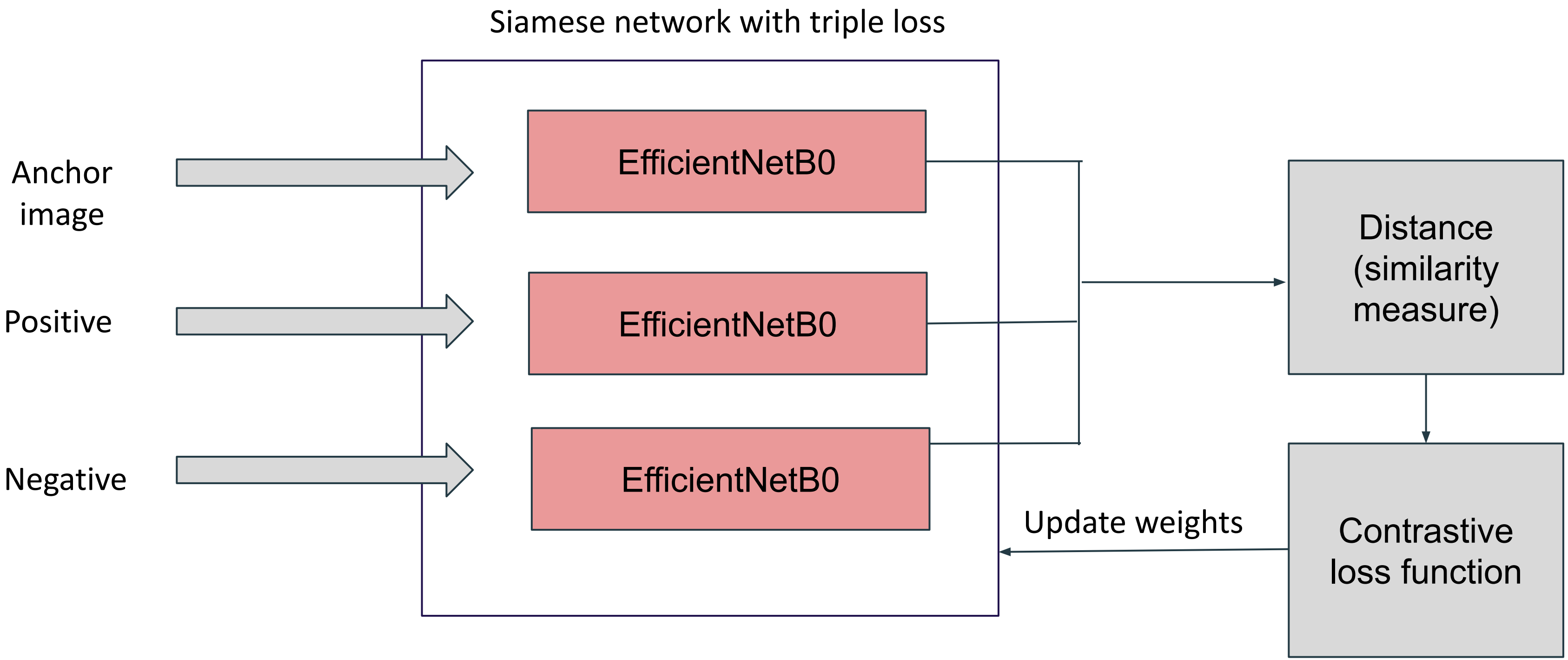
- VGG16
- RESNET-50
- INCEPTION V3
- INCEPTION RESNET V2
- EFFICIENT NET



Literature

Paper Title	Feature	Accuracy
Discriminative ensemble learning for few-shot chest x-ray diagnosis	Used autoencoders approach for few shot	76%
COVID-19 detection from scarce chest x-ray image data using few-shot deep learning approach	Used siamese	96.4%

Novelty of our proposed system



Observations

Architecture	Train	Test	No of epochs	Data
Siamese + CNN with batch normalization and dropout (Baseline)	90.37%	85%	50	80%
Siamese + Efficient Net	**99%**	**97%**	**8**	**80%**

References

1. Zhao, Defang et al. "Synthetic Medical Images Using F&BGAN for Improved Lung Nodules Classification by Multi-Scale VGG16." Symmetry 10 (2018): 519
2. Onishi, Yuya et al. "Automated Pulmonary Nodule Classification in Computed Tomography Images Using a Deep Convolutional Neural Network Trained by Generative Adversarial Networks." BioMed research international vol. 2019 6051939.
3. Albahli S. Efficient GAN-based Chest Radiographs (CXR) augmentation to diagnose coronavirus disease pneumonia. Int J Med Sci. 2020;17(10):1439-1448. Published 2020 Jun 6. doi:10.7150/ijms.46684
4. Tan L, Jing Y, Huo YT, Akin O. Lgan: Lung segmentation in CT scans using generative adversarial network. arXiv preprint arXiv:1901.03473, 2019.

References

5. F. Munawar, S. Azmat, T. Iqbal, C. Grönlund and H. Ali, "Segmentation of Lungs in Chest X-Ray Image Using Generative Adversarial Networks," in *IEEE Access*, vol. 8, pp. 153535-153545, 2020, doi: 10.1109/ACCESS.2020.3017915.
6. Chouhan V., Singh S. K., Khamparia A., Gupta D., Tiwari P., Moreira C., et al., "A Novel Transfer Learning-Based Approach for Pneumonia Detection in Chest X-ray Images," Applied Sciences, vol. 10, p. 559, 2020.
7. Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, and Andrew Y. Ng. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. arXiv:1711.05225 [cs, stat], November 2017. URL http://arxiv.org/abs/1711.05225. arXiv: 1711.05225.
8. Jadon, Shruti. (2021). COVID-19 detection from scarce chest x-ray image data using few-shot deep learning approach. 1. 10.1117/12.2581496.
9. Voulodimos A, Protopapadakis E, Katsamenis I, Doulamis A, Doulamis N. A Few-Shot U-Net Deep Learning Model for COVID-19 Infected Area Segmentation in CT Images. Sensors. 2021; 21(6):2215. https://doi.org/10.3390/s21062215

THANK YOU