NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY
[IJIERT] ISSN: 2394-3696 Website: ijiert.org
VOLUME 8, ISSUE 9, Sep. -2021

MACHINE LEARNING BASED CHEST X-RAY ABNORMALITIES DETECATION SYSTEM

Sandeep Singh
Department of computer science, Govt. Engineering College, Raipur, C.G.
sandeepsinghtekam007@gmail.com

Sushma Dewangan
Department of computer science, Govt. Engineering College, Raipur, C.G.
sush22dew@gmail.com

Nilam Singh
Department of computer science, Govt. Engineering College, Raipur, C.G.
Singhnilam83@gmail.com

ABSTRACT
When you have a broken arm, radiologists help save the day —and the bone. These doctors diagnose and treat medical conditions using imaging techniques like CT and PET scans, MRIs, and, of course, X -rays. Yet, as it happens when working with such a wide variety of medical tools, radiologists face many daily challenges, perhaps the most difficult being the chest radiograph. The interpretation of chest X-rays can lead to medical misdiagnosis, even for the best practicing doctor. Computer-aided detection and diagnosis systems (CADe/CADx) would help reduce the pressure on doctors at metropolitan hospitals and improve diagnostic quality in rural areas. Existing methods of interpreting chest X-ray images classify them into a list of findings. There is currently no specification of their locations on the image which sometimes leads to inexplicable results. A solution for localizing findings on chest X-ray images is needed for providing doctors with more meaningful diagnostic assistance.

Keywords: Radiology, Radiograph's, X-ray's, Vinlab, Yolo.

INTRODUCTION
We aim to localize and classify 14 types of thoracic abnormalities frome chest Radiogragh's. We'll work with a dataset consisting of 18,000scans that have been annotated by experienced radiologists. We will train our model with 15,000 independently-labelled images and we will collected via VinBigData's we based platform, Vinlab. Details on building the dataset can be found in the recent paper "VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations".

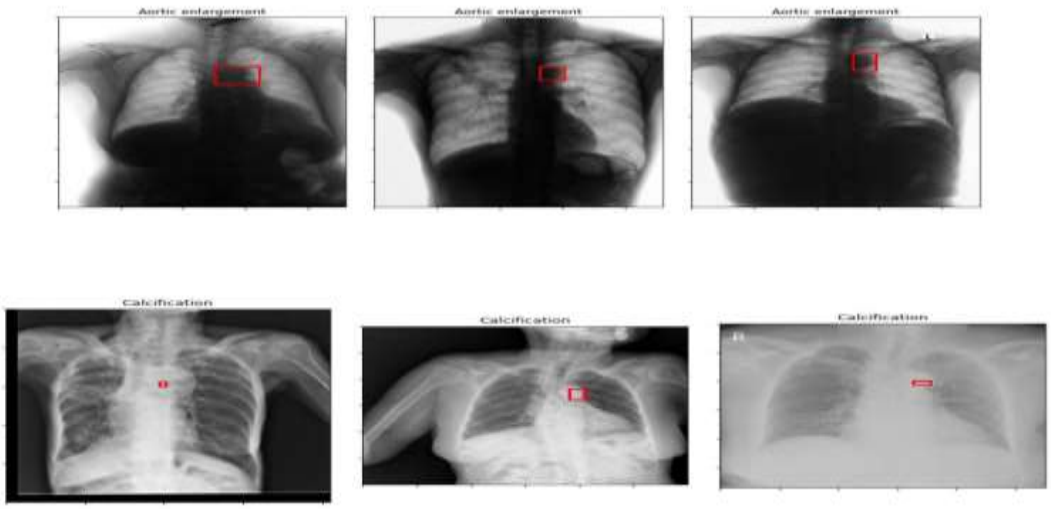

Figure 1: Minor Abnormalities find Through X-ray System

If successful, we'll help build what could be a valuable second opinion for radiologists. An automated system that could accurately identify and localize findings on chest radiographs would relieve the stress of busy doctors while also providing patients with
a more accurate diagnosis.

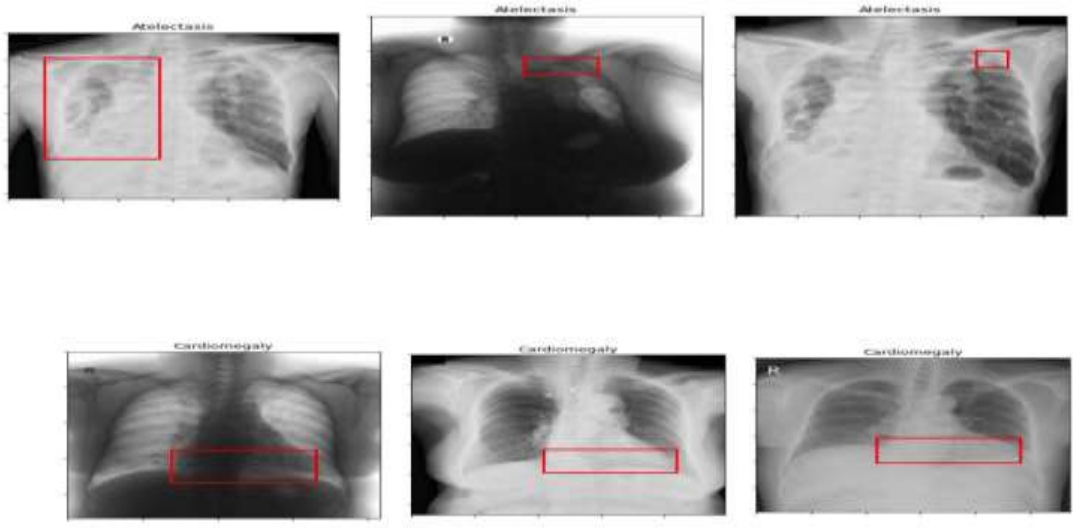

Figure 2: Major Abnormalities find Through X-ray System

Data Description (DD)

• We are classifying common thoracic lung diseases and localizing critical findings. This is an object detection and classification problem.

• For each test image, we'll be predicting a bounding box and class for all findings. If we predict that there are no findings, you should create a prediction of "14 1 0 0 1 1" (14 is the class ID for no finding, and this provides a one-pixel bounding box with a confidence of 1.0).

• The images are in DICOM format, which means they contain additional data that might be useful for visualizing and classifying.

• (DICOM): Digital Imaging and Communications in Medicine is the standard for the communication and management of medical imaging information and related data.

• By investigating the training dataset, we observed that the observations labelled 14 (no finding) largely outnumber those labelled with abnormalities. In order to avoid problems resulting from imbalance data, we removed rows that are labelled 14 (no finding). The goal of this project can be accomplished without using images with no finding. It is sufficient to use images with certain findings and bounding boxes to fit a model and make predictions. We then resized the input images to 512x512 pixels. Then we calculate and normalize x-mid, y-mid, width and height of the bounding boxes, and set up the labels in correspondence to 14 classes. We prepared data this way to fit into the YOLO model.

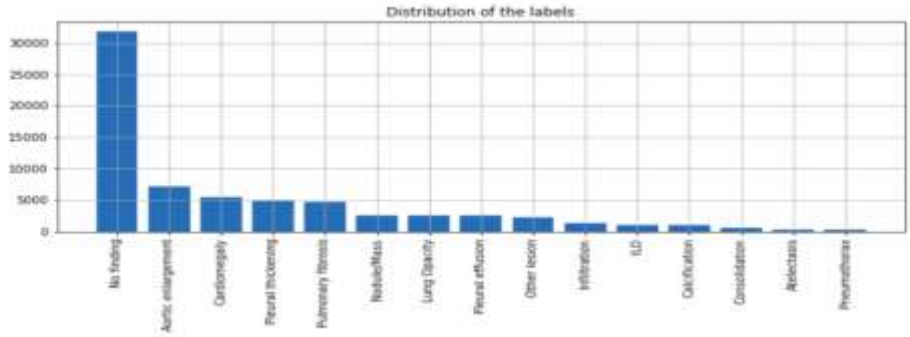

Figure 3: Distribution of Labels

We consider 14 critical radiographic findings as listed below (click for further information's):
0 - Aortic enlargement
1 - Atelectasis
2 - Calcification
3 - Cardiomegaly
4 - Consolidation
5 - ILD
6 - Infiltration
7 - Lung Opacity
8 - Nodule/Mass
9 - Other lesion
10 - Pleural effusion
11 - Pleural thickening
12 - Pneumothorax
13 - Pulmonary fibrosis

Data Access and Use:
Non-commercial, Research/Academic use only competitions are open to residents of the United States and worldwide, except that if you are a resident of Crimea, Cuba, Iran, Syria, North Korea, Sudan, or are subject to U.S. export controls or sanctions, you may not enter the Competition. Other local rules and regulations may apply to you, so please check your local laws to ensure that you are eligible to participate in skills-based competitions. The Competition Sponsor reserves the right to award alternative Prizes where needed to comply with local laws.

ENTRY IN THIS COMPETITION CONSTITUTES YOUR ACCEPTANCE OF THESE OFFICIAL COMPETITION RULES.
The Competition named above is a skills-based competition to promote and further the field of data science. You must register via the Competition Website to enter. Your competition submissions ("Submissions") must conform to the requirements stated on the Competition Website. Your Submissions will be scored based on the evaluation metric described on the Competition Website. Subject to compliance with the Competition Rules, Prizes, if any, will be awarded to participants with the best scores, based on the merits of the data science models submitted. See below for the complete Competition Rules.
We ask that you adhere to a code of honor and ethical behavior in your participation. This includes the strict prohibition against hand-labeling the test set, inclusive of using any hand-labelling of the test set to inform model training or selection. This competition is a unique opportunity to use machine learning to contribute to more accurate medical diagnosis and we appreciate your legitimate efforts to improve diagnostic accuracy.

Exploratory Data Analysis
• We investigated the training dataset by plotting the distribution of labels. Figure shows that Aortic enlargement (0) is the most common type of abnormality in our dataset. Other common types are Cardiomegaly (3), Pleural thickening (10), and Pulmonary fibrosis (13). Note that we removed the rows labelled no finding (14), which was initially the most common label in the training dataset.

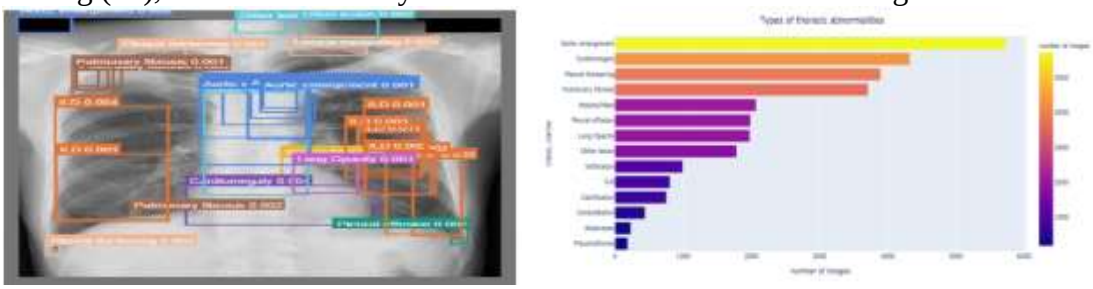

Figure 4: Training Dataset by Plotting the Distribution of Labels

Data Augumentation

• We augmented our dataset by flipping the images vertically or horizontally, rotating 90 degrees, adding noise, blurring, and changing brightness or contrast respectively. We didn't apply random cropping since it may even hinder experienced doctors to detect abnormalities. The probability for each augmentation step is a hyper-parameter that we can tune. We tried to populate our training set with 10%, 20%, 30%, 40%, and 50% augmented data respectively. We finalized on 20% since we didn't observe any further enhancement of mAP beyond 20%.

Evaluations

Our evaluation metric will be mAP@0.5 and mAP [0.5:0.95] for training and validation. The require's a mAP@0.4 for leader-board ranking, in other words our test set. We don't think this subtle difference in evaluation metric and actual loss function would affect our model selection, since mAPs at different IoUs are closely correlated to each other. We achieved a mAP@0.5 of 0.21 in our baseline model and boosted the performance to 0.30 with further pre-processing and data augmentation. We then select models from the Yolo family: Yolov5s, Yolov5x, Yolov5l, etc. and get a final mAP@0.5 of 0.34 with Yolov5x and further hyper-parameter tuning. Worth noticing, this 0.34 is derived from a test set that we split from the original training set rather than the leader-board test set.

	Baseline	Data Augmentation	Model Selection and Hyperparameter Tuning
mAP@0.5- local test set	0.21	0.30	0.34
mAP@[0.5:0.95]-local test set	0.08	0.13	0.16
mAP@0.4-leaderboard test set	0.24	NA	NA

Table1: Performance Improvement Progress

We once also implemented the train-devset configuration in our model selection process and didn't gather sufficient statistical evidence to support the alternative hypothesis that the distribution discrepancy due to data augmentation impairs prediction performance. Hence, we let the entire validation set follow the distribution of the test set.

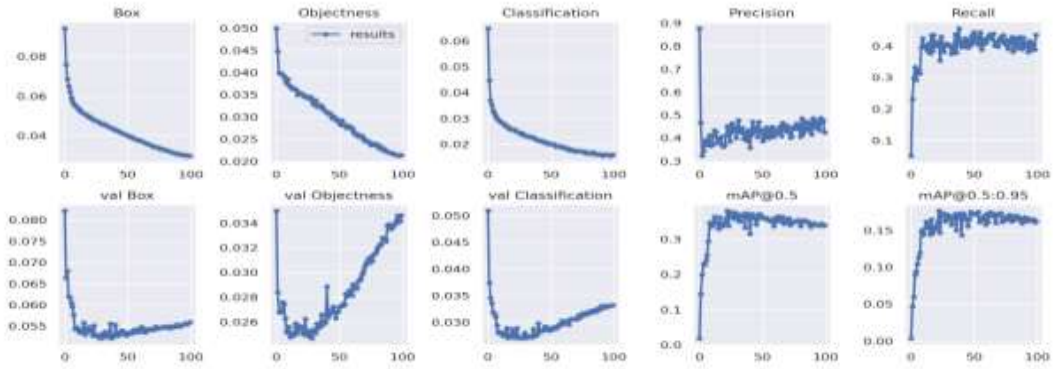

Figure 5: Traces of Metrics in Training Process

Error Analysis

After thorough analysis of error profile, we believe that there are three types of errors:

1. Imbalance category population as shown in Figure 6, our model predicts certain abnormalities reasonably well, while fails in others. Aortic Enlargement and Cardiomegaly are the two well-predicted abnormalities. Other Lesion, Infiltration, and Atelectasis are the worst performers. In fact, as shown in Figure 4, this contrast is extremely likely a result of an imbalanced number of samples in these categories. To solve this issue, we tried to oversample the minority cases. However, those minority cases often coexist with common abnormalities and therefore hinder our oversampling. In future, we may have to crop the minority cases first and then oversample. We may also consider cropping a few common cases to make all categories in the training set have exposure to cropped samples.
2. Indistinguishable background as shown in the last row of Figure 5, our model fails to spot many abnormalities and treats them as background. We don't have the necessary medical expertise to interpolate such errors, some of which may even be beyond top doctors' understanding. In future, we may ask medical experts for assistance to understand the underlying patterns of these errors.
3. Overlapping Objects Yolo families cannot handle cases that a large number of objects overlap due to the limitation of anchor boxes. Unfortunately, we are facing exactly this problem as shown in Figure 6. However, they doesn't allow us to use computationally expensive models, so for this project, we stick with Yolov5.

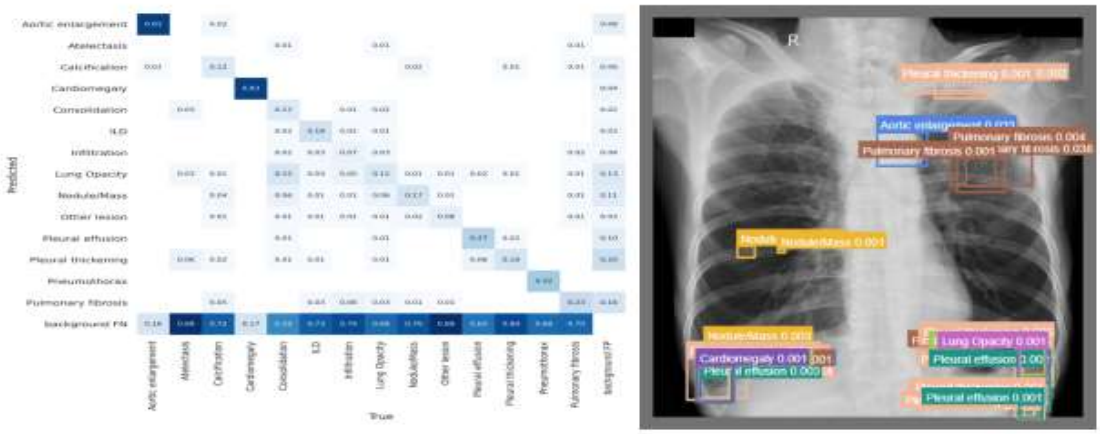

Figure 6: Conclusion Matrix & A training Sample with Exclusive Overlaps

Other Metrics

We reconfirm our previous observation that our model is especially predictive in Aortic Enlargement and Cardiomegaly through the lens of precision, recall, f1 score, and confidence

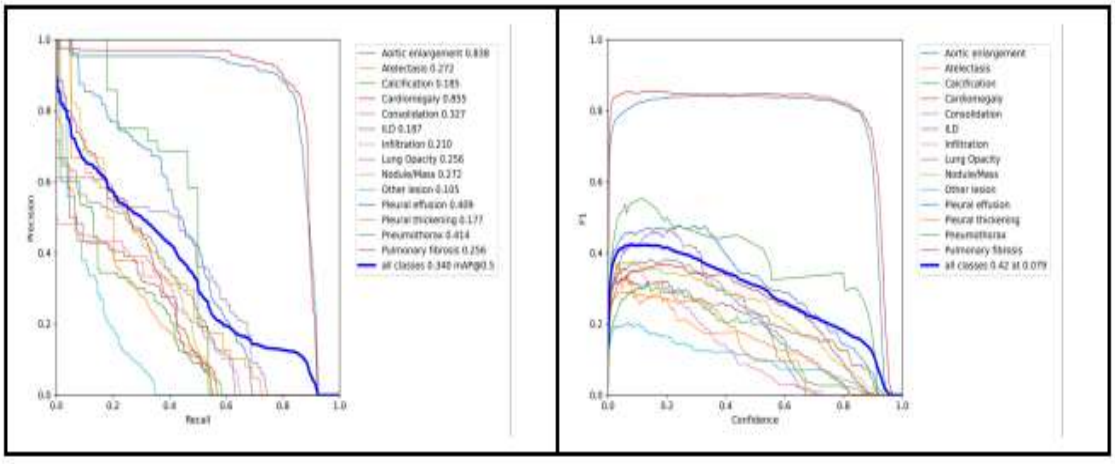

Figure 7: Precision Recall Curve & F1 Confidence Curve

RESULT AND DISCUSSION

By the help of many research studies and Algoritham's, preforming a project we can better understand the abnormalities of chest X-ray.

CONCLUSIONS

Probably we would like to conduct follow-up research to enhance performance by oversampling minority categories with cropping, further analyzing background related misclassification, and other models that handle overlapping well. We would also perform a more robust hyper parameter tuning, since the tuning process takes time longer than the scope of this project. This can be implemented in hospital systems, also this can assist doctors to readily rectify the problem and diagnose it.

REFERENCES

1) VinBigData Chest X-ray Abnormalities Detection | Kaggle.(2020)Kaggle. https://www.kaggle.com/c/vinbigdata-chest-xray-abnormalities-detection/overview
2) VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations. (2012). https://arxiv.org/pdf/2012.15029.pdf
3) CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. (2017). https://arxiv.org/abs/1711.05225
4) Automated abnormality classification of chest radiographs using deep convolutional neural networks. (2020). https://www.nature.com/articles/s41746-020-0273-z
5) Chest Radiograph Interpretation with Deep Learning Models: Assessment with Radiologist-adjudicated Reference Standards and Population-adjusted Evaluation. (2019). https://pubs.rsna.org/doi/10.1148/radiol.2019191293
6) YOLOv5 in PyTorch. (2020). Github. https://github.com/ultralytics/yolov5.