Chest X-ray abnormalities localization via ensemble of deep convolutional neural networks

Van-Tien Pham*, Cong-Minh Tran*, Stanley Zheng†, Tri-Minh Vu*, Shantanu Nath‡,
*Modelling and Simulation Centre, Viettel High Technology Industries Corporation, Vietnam
†Sir Winston Churchill & Athabasca University, Canada
‡Eutech Systems Ltd, Bangladesh
{tienpv13, minhtc3, minhvt8}@viettel.com.vn; szheng3@athabasca.edu; s.nath@eutechsystem.com;

***Abstract*—Convolutional neural network algorithms have been applied widely in chest X-ray interpretation thanks to the availability of high-quality datasets. Among them, VinDr-CXR is one of the latest public dataset including 18000 expert-annotated images labeled into 22 local position-specific abnormalities and 6 global suspected diseases. A proposed deep learning algorithms based on Faster-RCNN, Yolov5 and EfficientDet frameworks were developed and investigated in the task of multi-class clinical detection from chest radiography. The ground truth was defined by radiologist-adjudicated image review. Their performance was evaluated by the mean average precision (mAP 0.4), which can be accessed via Kaggle's server. The results shows the best performance belonging to ensembled detector model combined with EfficientNet as the classifier with the accuracy peak of 0.292. As a trade-off, ensembling detectors was much slower, which increases computing time by 3.75, 5 and 2.25 times compared to FasterRCNN, Yolov5 and EfficientDet, respectively. Overall, the classifiers shows constantly improvement on all detector models, which is highly recommended for further research. All of this aspects should be considered to address the real-world CXR diagnosis where the accuracy and computing cost are the most concerned.**

***Index Terms*—chest X-ray, abnormality detection, CNN, radiologist**

I. Introduction

Chest X-ray (CXR) diagnosis is a very important expertise enabling the ability to identify many types of diseases related to organs inside the chest area. The earlier and more accurate the diagnosis is, the more lives could be saved. Thus, diagnosis time and accuracy are the key factors for radiologists to concern . While the qualified-radiologist resources has not met the demand of this job [1], a computer-aided diagnosis (CAD) has been built and developed to support them in diagnosing common diseases or at least making decision faster and more accurate [2].

The rise of deep learning applications in medical image processing has been massive recently along with the availability of high-quality datasets with expert-generated annotations [3]. Deep convolutional neural network (CNN) has been improved remarkably to perform important medical applications at expert level [4]. In CXR diagnosis, some previous works showed good performance in interpreting CXRs with advanced CNN algorithm. For example, [5] proposed a deep learning algorithm to detect successfully abnormalities on CXRs but failed to categorize specific findings due to spectrum bias and lack of generalizability. Basically, these issues were partly solved by [6], using labeled data with radiologist-adjudicated reference. Their model was able to detect correctly pneumothorax, opacity, nodule or mass, and fracture from CXRs. For wider range of classification, [7] developed CheXNeXt to detect 14 different pathologies from chest x-ray focusing in thoracic diseases, then validated model with radiology experts. Although some models have claimed that their model reached expert-level performance [8], many aspects need to be considered for ready adoption in real-world applications [9]. Some research questioned the lack of data generalization accrossing institutions, the uncertainty of large-scale hand-annotation medical images and radiology-text meaning which reflecting medical nature [10]–[12].

To expand the hands-on applications in this field, this paper reported critical tasks of anomaly diagnosis for a wide range of abnormalities and diseases using different state-of-the-art neural networks. The VinDr-CXR datasets [3] was chosen with the training set of 15000 and a test set of 3000. Using web-based labeling tools developed by Vinlab, each image was annotated into 22 local abnormalities and 5 global diseases by board-certified radiologists. The main contributions of this work are listed as follows:

- An in-depth exploratory data analysis (EDA) was conducted on VinDr-CXR dataset to reflect its properties and then suggested the appropriate method.
- According to recent reviews [13], [14], there has been very few open pipeline for radiography abnormally detection. This work proposed a novel framework covering well-known object detectors. 2-class classifier was conducted on external dataset. Then, it was adopted in inference phase in order to reduce false positive and boost the performance of the whole pipeline. Cross validation strategy was applied in training phase while inference phase utilized many ensemble techniques [15], [16].
- Lastly, a comparative evaluation was analyzed in terms of accuracy and computing cost. The recommendation in CXR-diagnosis applications was highlighted in this section.

This paper was organized into 5 sections. After introduction, some related works was reviewed in section II. In section III, the proposed framework was described. Experimental results

then were analysed in section IV. Section V concluded and proposed recommendation for future works.

II. Related works

With high-quality dataset of CXRs such as CheXpert [17], Padchest [18], MIMIC-CXR [19] and recently VinDr-CXR [3], CNNs achitectures have seen a remarkable success in recent years. Among the most successful models, [20] classified large-scale dataset called ImageNet LSVRC-2010 into 1000 different classes with average error rate of 16.4%. Despite utilizing dropout regularization and non-saturating neurons, it still faced very high computing cost and time due to complex neural networks with 60 million parameters. To make multi-class classification tasks simpler as [21] showing outstanding performance, 2-class classifier was applied for external dataset in this research to quickly identify abnormal findings from normal CXRs. Further to multi-category classification for abnormal CXRs in inference phase, these algorithms have shown their drawbacks dealing with imbalanced classes in dataset as well as labeled errors during processing [22], [23]. Hence, this research utilized albumentation [24], a fast and flexible data augmentation technique to diversify the training and validation set by performing different transformations while keeping the same output labels. These transformation can be related to color, contrast, brightness, position and scale. This technique is very important due to the extreme imbalance of the VinDr-CXR dataset [3], which was presented in the next section. Moreover, dropout regularization was also applied to these models to reduce overfitting, which has been proved to be effective [25].

In this paper, we reported the deep-learning CNNs performance on the multi-category classification of CXRs. Compared to other architectures for COVID-19 binary classification from CXRs, while YOLO predictor was based to build successfully CAD to diagnose COVID-19 from other common diseases with accuracy of up to 97.4% [26], Faster-RCNN and EfficientNet based models reached the accuracy of 97.36% and 97%, respectively [27], [28]. For multi-class cases, YOLO-based model performed real-time up to 87% accuracy [29]. Based on these successful research, various CNN architectures such as YOLO [26], [30], Faster-RCNN [27], [31], and EfficientNet [28], [32] were trained and validated in this work, then evaluated by the test set which was labeled by the consensus of 5 radiologists [3]. For comparison, mean average precision was used as evaluation metric.

III. Proposed framework

This research proposed a framework as illustrated in Fig. 1. For model training, the flowchart includes 4 steps as shown in Fig. 2. First of all, the DICOM-format dataset is processed and resized into 1024x1024 images in png format. Secondly, external datasets such as ChestXray14 and Padchest are collected, then only global labels are filtered and unified into a classification dataset for training the 2-class classifier as described in III-B. Next, augmentations with Albumentations are adopted to diversify data to avoid overfitting and resolve the issue of imbalance dataset. The last step is to utilize the stratified k-fold cross validation during training detectors. Later, the final model zoo can be used for proposed framework. The following sections describe detail of each component including image augmentation, 2-class classifier, CXR abnormality detectors, validation strategy and ensemble technique. Finally, for testing, 3000 images of the test set are fed into 2-class classifier before applying different detectors to localize abnormalities. Then, Weighted Boxes Fusion (WBF) ensembles output from different techniques before validating final results.

A. Image augmentation

Image augmentation is used in computer vision tasks with the purpose of increasing the quality of trained models by diversifying label via creating new training samples from the existing data. Albumentations [24], a fast and flexible Python library, is currently the most popular augmentation library. Containing more than 70 different augmentations written by experts, it is widely used in industry, deep learning research, machine learning competitions, and open source projects. From the aforementioned EDA, we utilized Albumentation during training phase as illustrated in Fig. 3. Labels from multiple images were merged with many operators like blur, random contrast, RGB shift and channel shuffle.

B. 2-class classifier

The dataset is split up into 15 classes - 14 of which are abnormalities, and therefore must have bounding boxes placed, and one of which indicates no finding, and therefore, an absence in bounding boxes. This leads to an issue of excessive false positives - if an image indeed has no finding, we should not place bounding boxes. However, false positives are desired on images with abnormalities in order to optimize for the mAP metric. One possible way to mitigate this challenge is by separating the tasks of identifying images with no finding; we can use a classifier CNN to classify between images of no finding and images with abnormalities. Then, a detector is adopted to localize and classify into the remaining 14 abnormal classes. Experiments showed that Resnet50 [33] significantly accelerates the scores as well as reduces many false positive predictions on images with no finding. Thus, Resnet50 was adopted for its robustness. EfficientNet [34], which has shown promising results in other object classification tasks, is also integrated to this framework for ablation study purpose. As only global label is required for training classifier, all datasets in the data collection was used to train these CNNs. Diversity and huge amount samples from this training set makes this classifier very responsive and sensitive. In addition, we proposed a class-aware sampler to over-sample uncommon classes.

C. CXR abnormality detector

- ***FasterRCNN***: Originating from RCNN models , Faster-RCNN [35] is an notable representation of the two-stage object detection models. In these detectors, sparse region

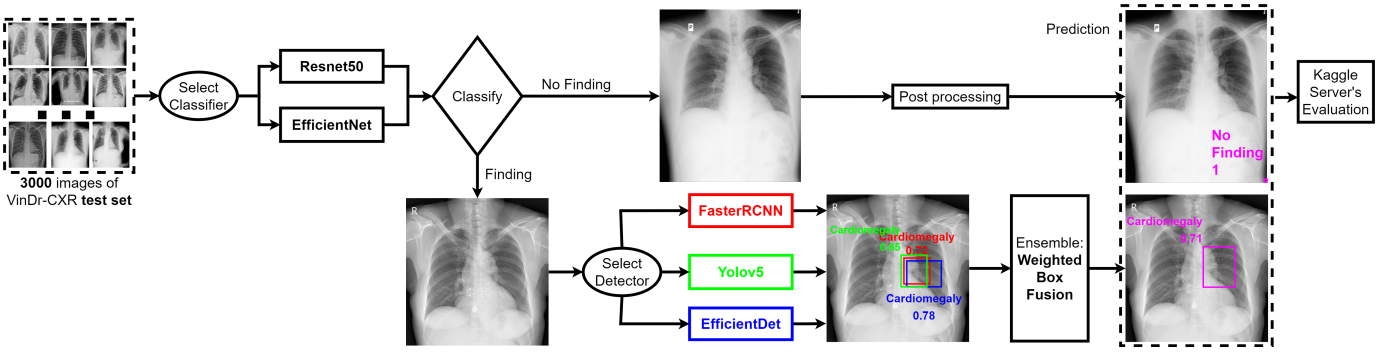


Fig. 1: Proposed framework for CXR abnormality detection.

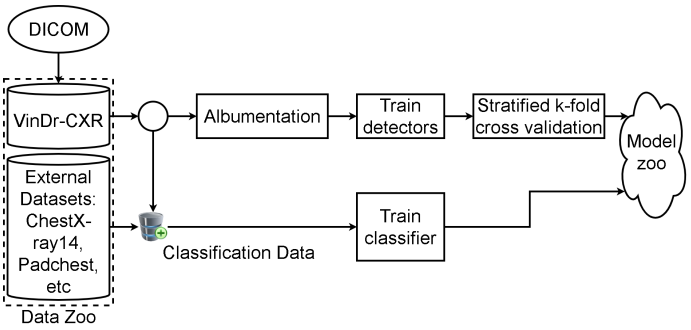


Fig. 2: Training flowchart for CXR abnormality detection.

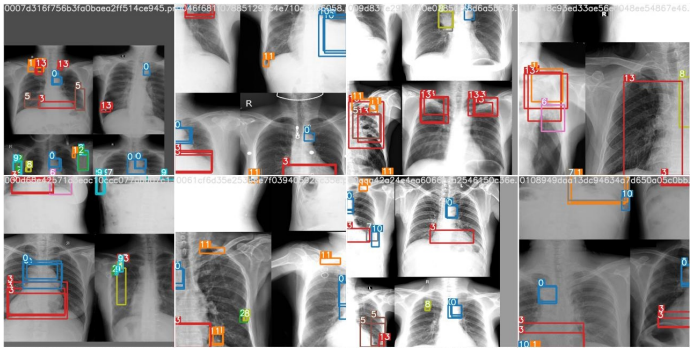


Fig. 3: Albumentation generation from VinDr-CXR dataset

proposals are generated in the first stage and then further regressed and classified in the second stage. RCNN utilizes Selective Search to generate proposals, then adopts CNN to extract features for training SVM classifier and bounding box regressor. FastRCNN extracts features for each proposal on a shared feature map by spatial pyramid pooling. It integrates the region proposal process into the deep convnet and makes the entire detector an end-to-end trainable model. In this paper, we employed FasterRCNN model pre-trained on COCO dataset and then fine-tune on CXR datasets via Detectron2 [36].

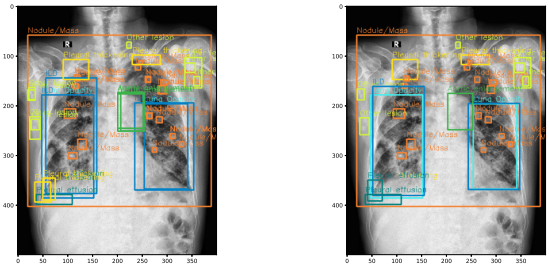
- ***Yolov5***: Single-stage detectors, such as YOLO and SSD treat object detection as a simple regression problem by taking an input image and learning the class probabilities and bounding box coordinates. Such models reach lower accuracy rates, but much faster than two-stage object detectors. We adopted YOLOv5 as the latest version from the YOLO model family [37]. It improves YOLOv4 with several bags of features and modules such as SiLU activation and Mosaic data augmentation to achieve robustness along with an impressive accuracy.
- ***EfficientDet***: EfficientDet [38] achieves the best performance in the fewest training epochs among object detection model architectures, making it a highly scalable architecture especially when operating with limited computing resources. The detector is the object detection version of EfficientNet, building on the its image classification tasks. It has been developed by (i) a weighted bi-directional feature pyramid network (BiFPN) with better accuracy and efficiency trade-offs, which allows easy and fast multiscale feature fusion; (ii) a compound scaling method that uniformly scales the resolution, depth, and width for all backbone, feature network, and box/class prediction networks at the same time . The network was delivered in a series of model sizes D0-D7. The model of choice for our work is EfficientDet-D7 which pretrained on the COCO dataset.

D. Validation strategy

We used group multiple label stratified k-folds, stratified on the class labels of each image, and grouped images based on its names. We evaluate 2-stage models together by combining predictions from the detection model along with the 2-class classifier to produce a unified set of labels, which were then optimized based on mAP 0.4. For our single stage detection, we used identical postprocessing, evaluating following post-processing.

E. Ensemble technique

The result of an object detection model is the location of an object with the confidence score of a class. As the predicted bounding box depends on various features of the class, the model generates many bounding boxes for a single class. Non-Maximum Suppression (NMS) is one of the techniques to overcome the problem where only a predicted box from the list of boxes is taken into consideration based on an IoU threshold.



(a) before WBF (b) after WBF

Fig. 4: Example of applying WBF ensemble technique.

TABLE I: Description of collected chest X-ray datasets.

Dataset	#Class	#Label	Annotation	Year
MC [39]	1	138	Classification	2014
SH [39]	1	662	Classification	2014
Indiana [40]	10	8121	Classification	2016
ChestX-ray14 [41]	14	112120	Classification	2017
CheXpert [17]	14	224316	Classification	2019
Padchest [18]	193	160868	Classification	2019
MIMIC-CXR [19]	14	377110	Classification	2019
JSRT [42]	1	247	Detection	2000
ChestX-ray8 [41]	8	108948	Detection	2017
VinDr-CXR [3]	15	18000	Detection	2020

NMS works well for single model prediction whereas for ensembling multiple models, Weighted Boxes Fusion (WBF) [15] shows better results compared to NMS and Soft-NMS [16]. Unlike NMS, WBF calculates the average of all predicted confidence scores and bounding boxes instead of eliminating extraneous boxes. Fig. 4 shows an example of ensemble of bounding boxes from 14-class detection.

subfig

IV. Experiments

A. Chest X-ray data collection

Tab. I summarizes CXR data collection from available public dataset. Among them, VinDr-CXR dataset is the latest one with 18,000 postero-anterior (PA) CXR scans in DICOM format, which were de-identified to protect patient privacy [3]. All images were labeled by a panel of experienced radiologists for the presence of 14 critical radiographic findings such as Atelectasis, Pneumothorax, etc. The label correlogram visualizes the dataset in a glimpse as shown in Fig. 5. The top-left subfigure shows clearly the imbalance among classes, especially between normal CXRs (class 1) and others (class 2-14). It is improved by albumentation augmentation technique as mentioned in III-A.

B. Evaluation protocol

All experiments were conducted on a Kaggle notebook using Tesla P100 16GB. We used the mean Average Precision (mAP) at IoU 0.4 since this is the most popular metric in object

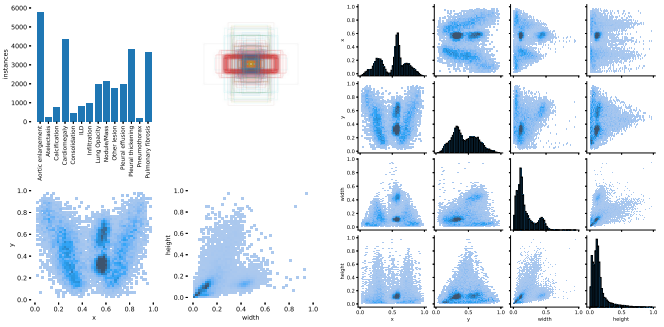


Fig. 5: Labels correlogram. Top-left histogram shows class frequency while the next to plot is a snapshot of labels in normalized height and width space. Diagonal histograms represent the distribution of each variable (x, y, width, height) of the finding's centroid and bounding box. Therelationship between each pair of variable is visualised via scatterplots.

detection domain. Submission was uploaded via Kaggle server and was evaluated by the competition's host.

C. Results and discussions

Tab. II shows comparison between methods integrated in the proposed framework on VinDr-CXR dataset. Overall, as classifiers are concatenated before detectors, the outcomes expands remarkably. In Fig. 6, the single model represented by the red line performs consistently worse than other concatenated methods regardless different detectors applied, though it gets accuracy peak of 0.272 in ensembling case. To highlight the importance of classifiers, for instance, the Yolov5 model's accuracy increases from 0.248 to 0.263 and 0.278 when Resnet50 and EfficientNet are respectively applied. This demonstrates that in the test dataset, many samples were found false positive by the detectors. The utilization of a classifier lessens these circumstances, subsequently diminishes the tension on the detection stage and increased the accuracy of the entire system. In particular, the classifiers greatly increase the accuracy of the system even in case of applying to a weak detector such as FasterRCNN. In this case, the mAP has increased by 17% and nearly 30% when applying Resnet50 and EfficientNet respectively. However, with a highly accurate detector like EfficientDet, the use of additional classifiers showed no significant improvement, only 0.011 and 0.004 for Resnet50 and EfficientNet respectively.

Another point that can be drawn from Figure 6 is that the EfficientNet classifier shows superiority over the Resnet50. On the same detector, EfficientNet usually gives a higher result than Resnet50. This finding is plausible and expected since many studies have shown the complexity and enhancement of the EfficientNet model compared to Resnet50 on several datasets [28], [32]. When combined with FasterRCNN, Yolov5, and Ensemble, EfficientNet outperformed Resnet50. However, in the EfficientDet case, the EfficientNet, with an accuracy of 0.273, proved to be inferior to the Resnet50's accuracy of 0.28. Because the classifier did not detect new

TABLE II: Evaluation of the proposed framework on VinDr-CXR dataset

Detector	Accuracy (mAP@0.4)			Performance		
	Single model	Resnet50	EfficientNet-B7	Speed (FPS)	GPU memory requirement (MB)	Training time (hour)
FasterRCNN	0.21	0.246	0.269	15	3291	**7**
YOLOv5	0.248	0.263	0.278	**20**	**2076**	9.5
EfficientDet-D7	**0.269**	**0.28**	0.273	9	3685	12
Ensemble	0.272	0.285	**0.292**	4	3685	30.5

functionality due to the replication of using EfficientNet as the backbone network in the EfficientDet model.

The experiments in this study indicate an obvious effect of an ensemble tactic, particularly the WBF in this case . Combining the individual results of each model together, the end result always increases in all three scenarios. The fusion of the EfficientNet classifier, led by single detectors, improves accuracy the most, rising from 0.278 to 0.292. For single detectors, the outcome is boosted from 0.269 to 0.272. This WBF technique has been widely used by most teams in this competition.

In term of single detection model precision, Tab. II evidently indicates that EfficientDet with a score of 0.256 generally outperforms the other detectors. Yolov5 ranked second with 0.248 while FasterRCNN proved quite weak compared to peers and attains only a fifth. Among those detectors, Faster-RCNN shows the worst results, with the best performance accuracy of 0.269 when combining with EfficientNet.

When assessing system performance, we only need to compare the time and memory taken when referencing on individual detectors regardless of whether or not classifiers are connected since the computational bottleneck mainly lies in the detection stage. Of course, the computational cost of the classifier does exist, but but it is minor in comparison to the detector. From Tab. II, the only single-stage detector, Yolov5, gives the fastest reference speed, 20 FPS. Deputy of two-stage detectors, FasterRCNN references a time of 15 FPS while EfficientDet is the slowest at about 9 FPS. The memory required when referencing the Yolov5 model is also impressive, just 2GB of GPU while its peers require over 3GB GPUs. The training time is fastest in the case of FasterRCNN with 7 hours and is the longest in the case of EfficientDet with up to 12 hours. In terms of computational cost, Yolov5 clearly strikes the optimal balance in terms of speed, memory requirement and required training time. One thing to keep in mind by using ensemble method is that we have to pay an expensive cost in terms of speed and training time. The reference speed is greatly reduced, only 4 FPS, because we need to aggregate the results separately from all single models, so it takes time to refer sequentially samples on each model. Similarly, the training time increases dramatically to more than 30 hours because all individual models need to be trained. Running the models in parallel is not feasible, so the required memory is calculated to be the maximum of the required memory in all cases, approximately 3,5GB GPU.

From the individual analysis mentioned above, we draw some suggestions about the selection of the appropriate

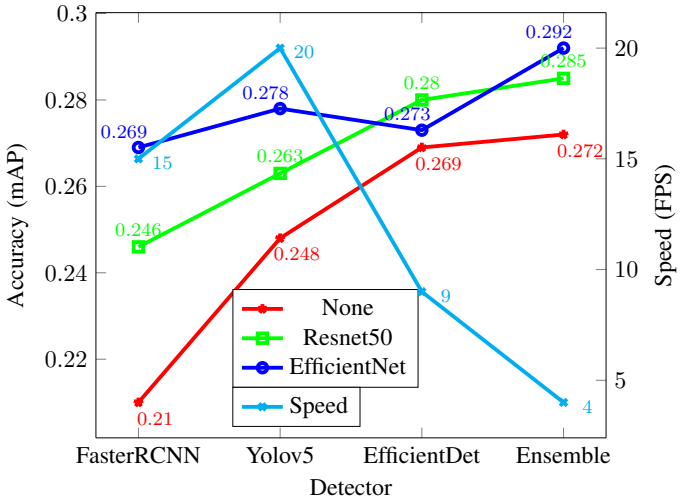


Fig. 6: Evaluation of combination of classifier and detector in the proposed framework.

method for practical applications. For tasks that require a high level of precision, using both a classifier and an ensemble technique is essential. We propose to train the data with all three detectors alongside with the EfficientNet classifier, and then use an ensemble technique in reference stage. Although there is a high cost, this method will be very effective in cases such as research labs, or cases of life-threatening illnesses. On the contrary, in the situation of a need for fast diagnostic rates for the purpose of basic screening to save costs, or when the number of samples to be predicted is large, the single model Yolov5 seems to be the most reasonable option. An additional combination of the EfficientNet classifier may be considered to improve accuracy at a small cost. Another option is to use EfficientDet for a slightly higher accuracy at a slower speed. Although the single model FasterRCNN proves inferior in terms of accuracy and cost, it is still possible to utilize this model into an ensemble technique.

V. Conclusions

In conclusion, this research highlighted the importance of solving the class imbalance in working with VinDr-CXR, one of the latest high-quality datasets. A proposed framework with a combination of 3 classifier options and 4 model detectors was tested and cross-validated with the test dataset. The 2-class classifier and ensemble technique for detector model was highly recommended for further study since their combination presented the best accuracy of 0.292. The computing time and system performance were also analyzed for a

well-rounded consideration when applying in different CXR-diagnosis situations. Code is at https://github.com/pvtien96/CXRAbnormalityLocalization.

REFERENCES

[1] N. Crisp and L. Chen, "Global supply of health professionals reply," *The New England journal of medicine*, vol. 370, pp. 950–7, 03 2014.

[2] H. H. Pham, T. T. Le, D. Q. Tran, D. T. Ngo, and H. Q. Nguyen, "Interpreting chest x-rays via cnns that exploit hierarchical disease dependencies and uncertainty labels," 2020.

[3] H. Q. Nguyen, K. Lam, L. T. Le, H. H. Pham, D. Q. Tran, D. B. Nguyen, D. D. Le, C. M. Pham, H. T. T. Tong, D. H. Dinh, C. D. Do, L. T. Doan, C. N. Nguyen, B. T. Nguyen, Q. V. Nguyen, A. D. Hoang, H. N. Phan, A. T. Nguyen, P. H. Ho, D. T. Ngo, N. T. Nguyen, N. T. Nguyen, M. Dao, and V. Vu, "Vindr-cxr: An open dataset of chest x-rays with radiologist's annotations," 2021.

[4] P. Moeskops, J. M. Wolterink, B. H. M. van der Velden, K. G. A. Gilhuijs, T. Leiner, M. A. Viergever, and I. Isgum, "Deep learning for multi-task medical image segmentation in multiple modalities," *CoRR*, vol. abs/1704.03379, 2017.

[5] R. Singh, M. Kalra, C. Nitiwarangkul, J. Patti, F. Homayounieh, A. Padole, P. Rao, P. Putha, V. Muse, A. Sharma, and S. Digumarthy, "Deep learning in chest radiography: Detection of findings and presence of change," *PLOS ONE*, vol. 13, p. e0204155, 10 2018.

[6] A. Majkowska, S. Mittal, and Steiner, "Chest radiograph interpretation with deep learning models: Assessment with radiologist-adjudicated reference standards and population-adjusted evaluation," *Radiology*, vol. 294, p. 191293, 12 2019.

[7] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Y. Ding, A. Bagul, C. Langlotz, K. S. Shpanskaya, M. P. Lungren, and A. Y. Ng, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *CoRR*, vol. abs/1711.05225, 2017.

[8] J. Nam, S. Park, E. J. Hwang, J. Lee, K.-N. Jin, K. Lim, T. Vu, J. Sohn, S. Hwang, J. M. Goo, and C. M. Park, "Development and validation of deep learning–based automatic detection algorithm for malignant pulmonary nodules on chest radiographs," *Radiology*, vol. 290, p. 180237, 09 2018.

[9] F. Cabitza, R. Rasoini, and G. F. Gensini, "Unintended Consequences of Machine Learning in Medicine," *JAMA*, vol. 318, no. 6, pp. 517–518, Aug. 2017.

[10] J. Zech, M. Badgeley, and M. Liu, "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study," *PLOS Medicine*, vol. 15, p. e1002683, 11 2018.

[11] H.-C. Shin, L. Lu, and R. Summers, *Natural Language Processing for Large-Scale Medical Image Analysis Using Deep Learning*, 12 2017, pp. 405–421.

[12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[13] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and future," *CoRR*, vol. abs/1704.06825, 2017.

[14] A. Majkowska, S. Mittal, D. Steiner, J. Reicher, S. McKinney, G. Duggan, K. Eswaran, and Chen, "Chest radiograph interpretation with deep learning models: Assessment with radiologist-adjudicated reference standards and population-adjusted evaluation," *Radiology*, vol. 294, p. 191293, 12 2019.

[15] R. Solovyev and W. Wang, "Weighted boxes fusion: ensembling boxes for object detection models," *CoRR*, vol. abs/1910.13302, 2019.

[16] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, "Improving object detection with one line of code," *CoRR*, vol. abs/1704.04503, 2017.

[17] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, and S. Ciurea-Ilcus, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," *CoRR*, vol. abs/1901.07031, 2019.

[18] A. Bustos, A. Pertusa, and J.-M. Salinas, "PadChest: A large chest x-ray image dataset with multi-label annotated reports," *Medical Image Analysis*, vol. 66, p. 101797, Dec. 2020, arXiv: 1901.07441.

[19] A. E. W. Johnson, T. J. Pollard, S. J. Berkowitz, and N. R. Greenbaum, "MIMIC-CXR: A large publicly available database of labeled chest radiographs," *CoRR*, vol. abs/1901.07042, 2019.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[21] Y.-X. Tang, Y.-B. Tang, Y. Peng, K. Yan, M. Bagheri, B. A. Redd, C. J. Brandon, Z. Lu, M. Han, J. Xiao *et al.*, "Automated abnormality classification of chest radiographs using deep convolutional neural networks," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–8, 2020.

[22] P. Rajpurkar, J. Irvin, R. L. Ball, and Zhu, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists," *PLOS Medicine*, vol. 15, no. 11, pp. 1–17, 2018, publisher: Public Library of Science.

[23] L. Oakden-Rayner, "Exploring large scale public medical image datasets," *arXiv:1907.12720 [cs, eess]*, Jul. 2019, arXiv: 1907.12720.

[24] A. Buslaev, A. Parinov, E. Khvedchenya, V. I. Iglovikov, and A. A. Kalinin, "Albumentations: fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, Feb. 2020, arXiv: 1809.06839.

[25] A. Mittal, D. Kumar, M. Mittal, and Saba, "Detecting pneumonia using convolutions and dynamic capsule routing for chest x-ray images," *Sensors*, vol. 20, no. 4, 2020.

[26] M. A. Al-antari, C.-H. Hua, J. Bang, and S. Lee, ""Fast deep learning computer-aided diagnosis of COVID-19 based on digital chest x-ray images"," *Applied Intelligence*, Nov. 2020.

[27] K. Shibly, S. Dey, M. Tahzib-Ul-Islam, and M. M. Rahman, "Covid faster r–cnn: A novel framework to diagnose novel coronavirus disease (covid-19) in x-ray images," *Informatics in Medicine Unlocked*, vol. 20, p. 100405, 08 2020.

[28] N. K. Chowdhury, M. A. Kabir, M. M. Rahman, and N. Rezoana, "Ecovnet: An ensemble of deep convolutional neural networks based on efficientnet to detect covid-19 from chest x-rays," 2020.

[29] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of covid-19 cases using deep neural networks with x-ray images," *Computers in Biology and Medicine*, vol. 121, p. 103792, 2020.

[30] Y. Kim, Y. Cho, and Wu, "Short-term reproducibility of pulmonary nodule and mass detection in chest radiographs: Comparison among radiologists and four different computer-aided detections with convolutional neural net," *Scientific Reports*, vol. 9, p. 18738, 12 2019.

[31] T. Griffin, Y. Cao, B. Liu, and M. J. Brunette, "Object Detection and Segmentation in Chest X-rays for Tuberculosis Screening," in *2020 Second International Conference on Transdisciplinary AI (TransAI)*, Sep. 2020, pp. 34–42.

[32] E. Luz, P. Silva, R. Silva, L. Silva, J. Guimarães, G. Miozzo, G. Moreira, and D. Menotti, "Towards an effective and efficient deep learning model for covid-19 patterns detection in x-ray images," *Research on Biomedical Engineering*, Apr 2021.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[34] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *CoRR*, vol. abs/1905.11946, 2019.

[35] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *CoRR*, vol. abs/1506.01497, 2015.

[36] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2," https://github.com/facebookresearch/detectron2, 2019.

[37] G. Jocher, A. Stoken, and J. Borovec, "ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations," Apr. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.4679653

[38] M. Tan, R. Pang, and Q. Le, "Efficientdet: Scalable and efficient object detection. arxiv 2019," *arXiv preprint arXiv:1911.09070*.

[39] S. Jaeger, S. Candemir, S. Antani, Y.-X. Wáng, P.-X. Lu, and G. Thoma, "Two public chest x-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative imaging in medicine and surgery*, vol. 4, pp. 475–7, 12 2014.

[40] D. Demner-Fushman, M. Kohli, M. Rosenman, S. Shooshan, L. Rodriguez, S. Antani, G. Thoma, and C. Mcdonald, "Preparing a collection of radiology examinations for distribution and retrieval," *Journal of the American Medical Informatics Association : JAMIA*, vol. 23, 07 2015.

[41] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," *CoRR*, vol. abs/1705.02315, 2017.

[42] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, and Kobayashi, "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *AJR. American journal of roentgenology*, vol. 174, pp. 71–4, 02 2000.