PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

# Assessing an AI-based smart imagery framing and truthing (SIFT) system to assist radiologists annotating lung abnormalities on chest x-ray images for development of deep learning models

Lin Guo, Kunlei Hong, Ziqi Zhang, Bin Zheng, Stefan Jaeger, et al.

Lin Guo, Kunlei Hong, Ziqi Zhang, Bin Zheng, Stefan Jaeger, Jordan Fuhrman, Hui Li, Maryellen Giger, Andrei Gabrielian, Alex Rosenthal, Darrell E. Hurt, Ziv Yaniv, Y. M. Fleming Lure, "Assessing an AI-based smart imagery framing and truthing (SIFT) system to assist radiologists annotating lung abnormalities on chest x-ray images for development of deep learning models," Proc. SPIE 12465, Medical Imaging 2023: Computer-Aided Diagnosis, 124650R (7 April 2023); doi: 10.1117/12.2653826

SPIE.

Event: SPIE Medical Imaging, 2023, San Diego, California, United States

# Assessing an AI-based Smart Imagery Framing and Truthing (SIFT) system to assist radiologists annotating lung abnormalities on chest X-ray images for development of deep learning models

Lin Guo[*a], Kunlei Hong[*a], Ziqi Zhang[*b], Bin Zheng[c], Stefan Jaeger[d], Jordan Fuhrman[e], Hui Li[e], Maryellen Giger[e], Andrei Gabrielian[f], Alex Rosenthal[f], Darrell E. Hurt[f], Ziv Yaniv[f], Y.M. Fleming Lure[#g]

[a]Shenzhen Zhying Medical Imaging Co., Ltd, Shenzhen, China; [b]Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, Shenzhen, China; [c]School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK, USA; [d]National Library of Medicine, National Institute of Health, Bethesda, MD, USA; [e]University of Chicago, Chicago, IL, USA; [f]Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institute of Health, Bethesda, MD, USA; [g]MS Technologies Corp, Rockville, Maryland, USA

[*]Lin Guo, Kunlei Hong and Ziqi Zhang contributed equally to this work.

[#]Corresponding author: Y.M. Fleming Lure, Email: f.lure@hotmail.com.

## ABSTRACT

To assess a Smart Imagery Framing and Truthing (SIFT) system in automatically labeling and annotating chest X-ray (CXR) images with multiple diseases as an assist to radiologists on multi-disease CXRs. SIFT system was developed by integrating a convolutional neural network based-augmented MaskR-CNN and a multi-layer perceptron neural network. It is trained with images containing 307,415 ROIs representing 69 different abnormalities and 67,071 normal CXRs. SIFT automatically labels ROIs with a specific type of abnormality, annotates fine-grained boundary, gives confidence score, and recommends other possible types of abnormality. An independent set of 178 CXRs containing 272 ROIs depicting five different abnormalities including pulmonary tuberculosis, pulmonary nodule, pneumonia, COVID-19, and fibrogenesis was used to evaluate radiologists' performance based on three radiologists in a double-blinded study. The radiologist first manually annotated each ROI without SIFT. Two weeks later, the radiologist annotated the same ROIs with SIFT aid to generate final results. Evaluation of consistency, efficiency and accuracy for radiologists with and without SIFT was conducted. After using SIFT, radiologists accept 93% SIFT annotated area, and variation across annotated area reduce by 28.23%. Inter-observer variation improves by 25.27% on averaged IOU. The consensus true positive rate increases by 5.00% ($p$=0.16), and false positive rate decreases by 27.70% ($p$<0.001). The radiologist's time to annotate these cases decreases by 42.30%. Performance in labelling abnormalities statistically remains the same. Independent observer study showed that SIFT is a promising step toward improving the consistency and efficiency of annotation, which is important for improving clinical X-ray diagnostic and monitoring efficiency.

**Key Words:** Chest radiograph; deep learning; image annotation; observer performance study

## 1. INTRODUCTION

The advancement of Deep Learning (DL) has greatly facilitated research on diagnostic radiology decision support, which relies on updated DL techniques, available training data, and rapidly growing computation capability[1-3]. Advances in

Medical Imaging 2023: Computer-Aided Diagnosis, edited by Khan M. Iftekharuddin, Weijie Chen, Proc. of SPIE Vol. 12465, 124650R · © 2023 SPIE
1605-7422 · doi: 10.1117/12.2653826

clinical decision support research provide achievements in different aspects, such as disease detection[4-7], disease progression prediction[8], abnormalities localization[9] and workflow improvement.

In general, to further boost the efficiency and generalization of the methods used in those clinical decision support, the most effective approach is to get access to large amount of high-quality image data[3, 10, 11]. High-quality image data requires accurate labelling of diseases and annotated boundary of the region of interests (ROIs). To ensure clinical generalizability of the model, it is desirable to have a large dataset with a variety of high-quality images from multiple institutions and different geographic regions[12]. However, cataloguing large datasets can be challenging due to their volume, limited radiologist availability, and time-consuming annotation processes. Collecting data for rare diseases is especially difficult. Additionally, many complexities are introduced in the data de-identification process to comply with patient privacy rules, institutional review board requirements, and local ethical committee protocols[13]. If training data are limited, deep learning–based models may suffer from overfitting, which results in poor generalizability.

Several studies work on this challenge by data collection, data annotation and label fixing. Yang et al.[14] specifically collected database many times consists of other abnormalities. Accurately labelling these comorbidities can increase the classification for deep learning. "Finer-grained annotation" ("Pixel-wise annotation"), instead of bonding box and ellipse-wise annotation, can not only improve the performance of deep learning models, but help deep learning model extract more accurate information. Fixing annotation mistakes can easily lead to a 10-20% improvement of the model and is considerably less effort than annotating extra data from scratch. In addition, adding new data will not help the accuracy of the model if the existing files still contain mistakes. It requires much more additional annotated training data to the neural network to get a similar accuracy improvement compared to just fixing the original dataset.

Therefore, sufficient annotation from radiologists is critical for training deep learning models to help the diagnostic radiology decision. However, about 25% of radiologists do not agree with other radiologists' diagnoses, and 30% do not agree with their own previous decisions[15]. The whole procedure of data annotation is very time consuming and tedious. Generally, in order to use deep learning to train networks to perform either of detection and segmentation tasks, one requires a fair amount of annotated images that mirrors the desired output: bounding boxes in the case of detection, and pixel-level masks in the case of segmentation. However, the state-of-the-art DL perform feature extraction, segmentation of ROI, then classification of each ROI in sequence. Accurate and consistent segmentation in training images will warrant the robustness and accuracy of DL classification than bonding box[16].

There are also some open-source graphical annotation tools for radiologists including labelme and it.snap. With available pathology report and a suite of templates (oval, circle, square, polygon, rectangular, triangle, square), radiologists manually label the locations of abnormality and select the best match template and manually place template on the center of the labeled abnormality. Additionally, some researchers propose automated annotation tools to improve the efficiency. MarkIT utilizes 1000 CXR with multi-label classifications by techniques of AI and blockchain to enable crowdsourcing and data exchange without segmentation to generate boundary. Its UI only changes brightness and contrast and the confidence is determined by annotators. Currently, automated annotation is less accurate than manual annotation, and therefore still needs expert validation to provide an error-free annotated dataset. In the existing methods, radiologists need to determine the center of the identified abnormalities and the best matched template. Algorithm determines the boundary of abnormality from the selected template in real time. Also, radiologists interact with algorithm which need to compute boundary in real-time.

In this work, we propose an AI-based Smart Imagery Framing and Truthing (SIFT) System which is designed to generate high-quality annotated cases to train machine learning and artificial intelligence. We design an experiment that the radiologist first manually annotated each ROI without SIFT and annotated the same ROIs with the assistance of SIFT to generate final results two weeks later. Evaluation of consistency, efficiency and accuracy for radiologists with and without SIFT was conducted.

## 2. METHODOLOGY

### 2.1 Material and Methods

#### 2.1.1. SIFT System

Smart Imagery Framing and Truthing (SIFT) System is designed to generate high-quality annotated cases to train machine learning and artificial intelligence. The annotated cases are not for the diagnosis and detection of diseases by physicians. It can automatically label 68 different diseases and annotate their boundary locations. Fig .1 displays our developed Expert-

Driven Fast Fine-Grainer Annotation & Training Process of SIFT system.

SIFT consists of graphic user interface (GUI), Network Connection and AI engine. AI engine (MOM ClaSeg algorithm) can automatically label the abnormality, its boundary, and confidence score, and other recommended abnormality. GUI will display AI engine labelled/annotated abnormality and allow annotators to manually edit the AI-generated results to generate final annotation. And the Network Connection is DICOM PACS connection, which support the data transmission between SIFT system and hospitals or other databases. Using SIFT system, annotators can label different AI predicted diseases and edit boundary. It is fully compliant with international standard (90% MedDRA coding / ICD 13 / RedLex and 10% Radiopaedia / PubMed). NO FDA Clearance is Needed because it is only intended to be used in research to generate data to train and validate the ML/AI system.

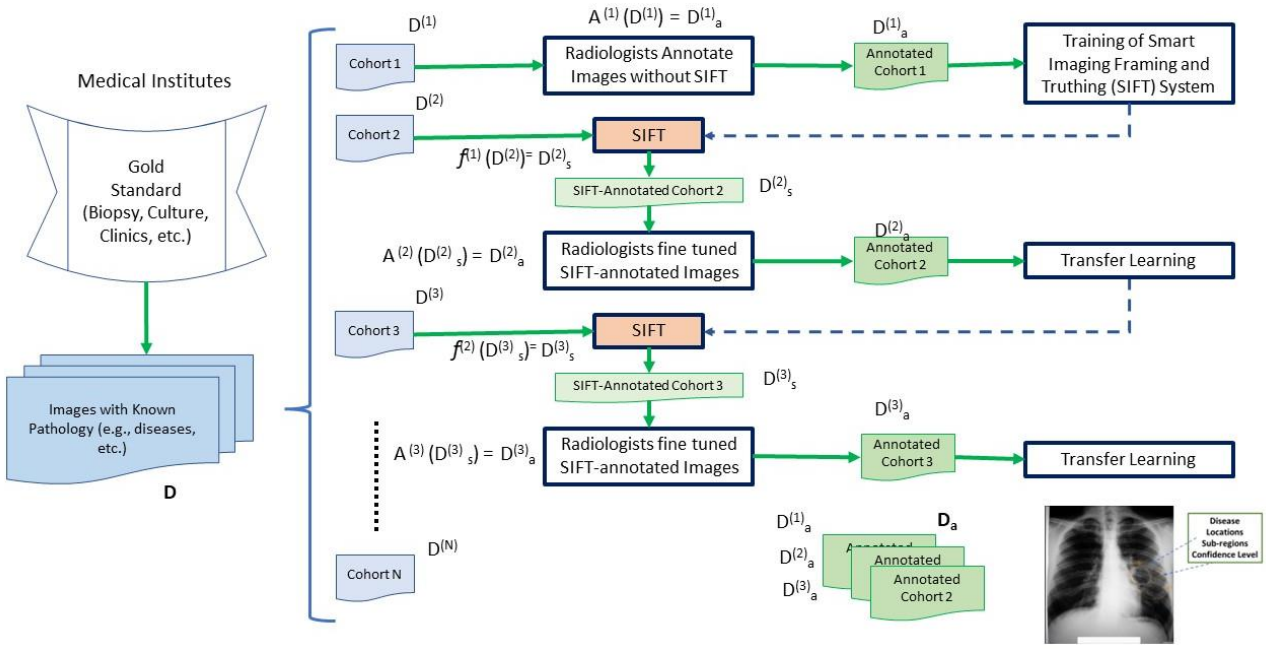



**Figure 1.** Expert-Driven Fast Fine-Grainer Annotation & Training Process

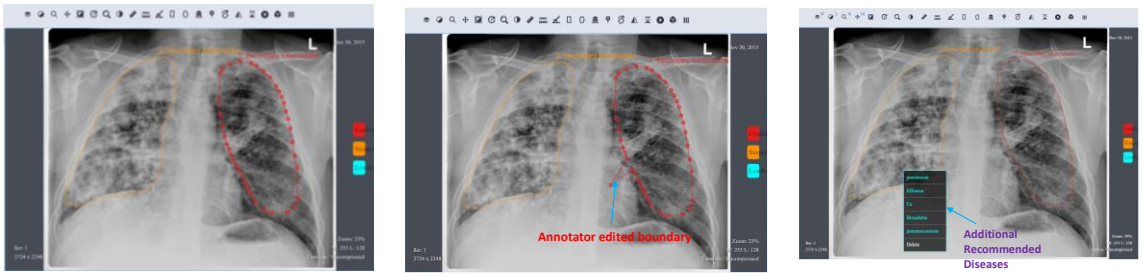
### 2.1.2 Training of AI Engine

SIFT system was developed by integrating a convolutional neural network based-augmented MaskR-CNN and a multi-layer perceptron neural network. It is trained with images containing 307,415 ROIs representing 69 different abnormalities and 67,071 normal CXRs[17]. It automatically labels ROIs with a specific type of abnormality, annotates fine-grained boundary, gives confidence score, and recommends other possible types of abnormality.

In order to reduce the risk of overfitting and determine the optimal training epochs and connection weights to build above transfer learning networks, we divided the training dataset using a ratio of 0.85 to 0.15 in all different image categories including normal images and 65 different types of abnormalities. Note that the term "normal image" is used to indicate that there is no finding in the CXR by radiologists. Thus, for each image category, 85% of CXR images are used to train the networks and 15% of CXR images are used to validate the network performance for different epochs and weights.

### 2.1.3 Graphic User Interface

The GUI displays AI engine labelled/annotated abnormality and confidence using Polyline annotation style, as shown in Fig 2.a. SIFT only displays portion of boundary points such that it allows easier editing by annotators, which allow annotators to manually edit the AI-generated results to generate final annotation (Figure 2b) and allow annotators to select other possible abnormalities and their corresponding boundaries (Figure 2c).




**(a)** **(b)** **(c)**

**Figure. 2** Graphic User Interface

### 2.1.4 Network Connection

DICOM viewer and annotation tools were implemented, includes features for image loading, parsing, decoding, and tools commonly encountered in DICOM viewers. Our platform is capable of fetching images from any vendor-neutral DICOM storage. We implemented a connection with both standard Picture Archiving and Communication System (PACS) systems. The output format can be JSON, Excel, or bmp.

### 2.2 Experimental Design

As shown in Table 1, an independent set of 178 CXRs containing 272 ROIs depicting five different abnormalities including pulmonary tuberculosis, pulmonary nodule, pneumonia, COVID-19, and fibrogenesis was used to evaluate radiologists' performance based on three radiologists in a double-blinded study. The dataset included 178 patients with 44.95% male, 22.47% female and 32.58% unknown (average age of 47±20 years).

Two expert radiologists with more than 30 years of experience serve as gold standard to determine the location of abnormality based on the pathology/diagnostic reports. Three study radiologists with 5-10 years of experience, used to annotate tuberculosis (TB) for several area on CXRs for at least 3 years, label the abnormality and annotate its location in the current study. Their observer performance and preference are studied and compared. The study radiologists first manually annotated each ROI without SIFT. Two weeks later, the radiologists annotated the same ROIs with SIFT aid to generate final results. The labelled abnormality, confidence for each abnormality (confidence score ranging from 0 to 5, where 0 is the lowest and 5 is the highest), boundary of labelled abnormality and labeling time for each case are recorded. Evaluation of consistency, efficiency and accuracy for radiologists with and without SIFT was conducted.

**Table 1.** Data description of testing images and regions of interest (ROI)

| Abnormality Types | Data Source | No. of Images | No. of ROIs |
|---|---|---|---|
| Pulmonary tuberculosis | NIAID public dataset | 44 | 50 |
| Pulmonary nodule | In-house | 38 | 43 |
| Pneumonia | In-house | 31 | 42 |
| COVID-19 GGO | BrixIA Public dataset | 32 | 63 |
| Fibrogenesis | In-house | 33 | 74 |
| **Total** | | **178** | **272** |

## 3. RESULTS

### 3.1 Evaluation of consistency for the same Types and Same Area of Abnormalities

### 3.1.1 Effect of SIFT on radiologists' decision on annotated area

We first evaluate the effect of SIFT on radiologists' decision on annotated pixel area. In Fig.3, it is shown the examples of annotated images for TB respectively from radiologist only, SIFT only, and radiologist using SIFT. It can be observed that radiologists changed the annotation and accepted the area resulted from SIFT after using and being assisted by SIFT. Fig. 4 demonstrates it more obviously by compare the overlapped images.

From the statistics (Table 2), radiologists accept 93% SIFT annotated area (93%=100% -7%). Radiologist's variation (STDEV) to annotate an area reduce by 28.23% after using SIFT. Without SIFT, radiologist annotated abnormality area is 29% larger than SIFT annotated area.

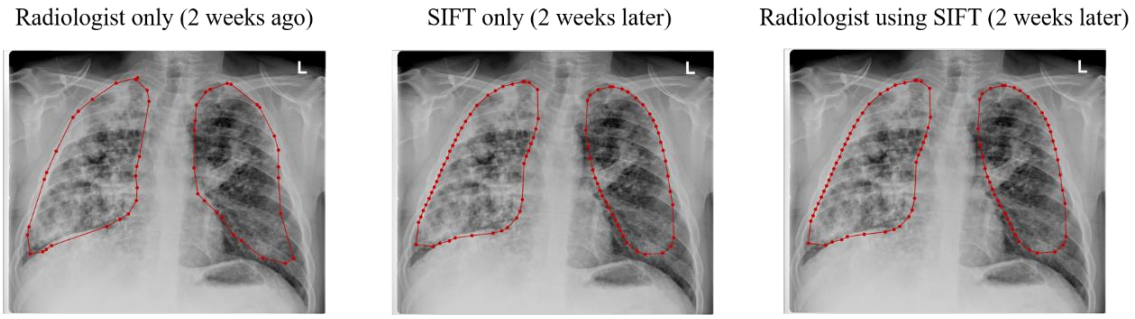



**Figure.3** Example of annotated images for TB with radiologist only (left), SIFT only (middle), and radiologist using SIFT (right).

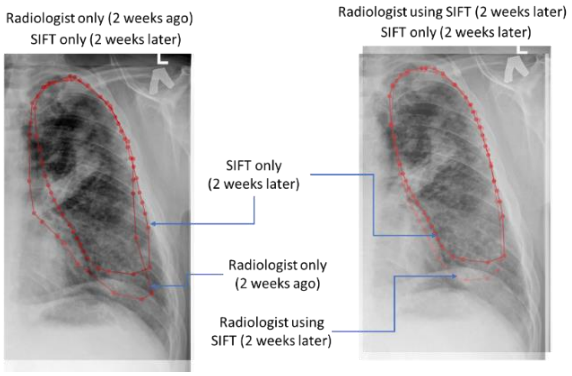



**Figure 4.** Example of overlapped images for TB with radiologist only (left), SIFT only (middle), and radiologist using SIFT (right).

**Table 2.** Effect of SIFT on Radiologists' Annotated Area

| | Radiologists only (2 weeks ago) | SIFT Only | Radiologists with SIFT Aid (2 weeks later) | Area Difference | | | Area Difference (%) | |
|---|---|---|---|---|---|---|---|---|
| | Area (B1) | Area (B2) | Area (B3) | B2-B1 | B3-B2 | B3-B1 | (B2-B1)/B1 | B3-B2)/B2 |
| Radiologist 1 | 104,278,048 | 72,891,095 | 75,944,200 | -31,386,953 | 3,053,105 | -28,333,849 | -30% | **4%** |
| Radiologist 2 | 109,229,920 | | 84,560,522 | -36,338,825 | 11,669,427 | -24,669,398 | -33% | **16%** |
| Radiologist 3 | 92,734,598 | | 72,834,616 | -19,843,503 | -56,480 | -19,899,983 | -21% | **0%** |
| Average Radiologist | 102,080,855 | | 77,779,779 | -29,189,760 | 4,888,684 | -24,301,076 | -29% | **7%** |
| Standard Deviation | 8,464,316.28 | | 6,074,638.38 | - | - | 2,389,677.90 | - | **-28.23%** |

### 3.1.2 Effect of SIFT on consensus area

As shown in Fig.4, we also conduct an analysis on five abnormalities based on intersection of union (IoU) Coefficient. The IoU can be calculated as:

$$IoU = \frac{Area\ of\ overlap}{Area\ of\ union} \qquad (1)$$

Where area of overlap and area of union are the overlapping area and union area between two times annotation respectively.

Consensus in annotated overlapped area improves by 25.27% on average in IoU. Note according literatures[15], inter-observer variation for two radiologists can be as high as 25%. The Inter-observer variation in our study reduced by 39.62%. Improvement for radiologists before and after use of SIFT is the lowest for TB because these radiologists are very familiar with annotating TB.

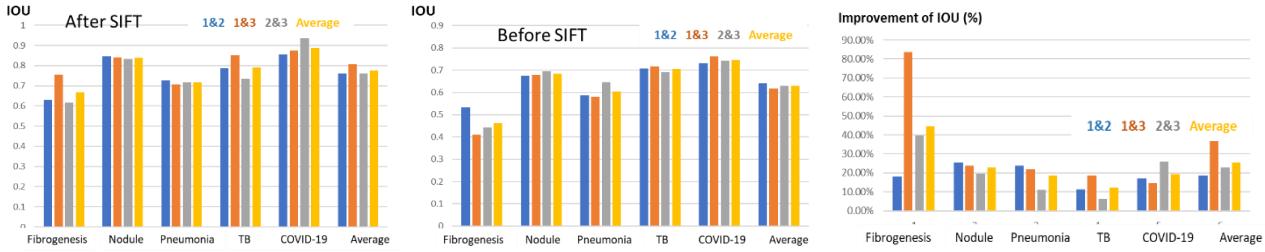



**Figure 4.** IOU marked by two radiologists before and after SIFT. Inter-observer variation in annotating area improves by 25.27% on average in IOU. IOU improvement for radiologists before and after use of SIFT is the lowest for TB because these radiologists are very familiar with annotating TB.

### 3.1.3 Comparison of labelled abnormalities (total ROI)

Accuracy of labelled ROI is based on the accuracy of AI-predicted ROI in terms of its location and class of abnormality. Total ROI can be calculated as:

$$\text{Total ROI} = \text{TP ROI} + \text{FP ROI}$$

Where TP ROI and FP ROI are True Positive ROI and False Positive ROI, respectively.

As shown in Table 3, Total ROI decreases by 6.93% after use of SIFT. Consensus ROI decreases by 3.10%. No consensus ROI decreases by 18.18%. The consensus true positive rate increases by 5.0% after use of SIFT ($P$=0.16). The consensus false positive rate decreases by 27.7% after use of SIFT ($P$<0.001).

**Table 3.** Radiologists' consensus in labelling of ROI with and without SIFT

| | True Positive ROI | | | False Positive ROI | | | Total ROI | | |
|---|---|---|---|---|---|---|---|---|---|
| ROI labelled by Radiologists | Consensus | No Consensus | Sub-total | Consensus | No Consensus | Sub-total | Consensus | No Consensus | Total |
| Without SIFT assistance | 201 | 26 | 227 | 65 | 78 | 143 | 266 | 104 | 370 |
| With SIFT assistance | 211 | 28 | 239 | 47 | 60 | 107 | 258 | 88 | 346 |
| Change of labelled ROI | 5.00% | 7.69% | 5.29% | -27.70% | -23.08% | -25.17% | -3.10% | -18.18% | -6.93% |
| *P* value | | *P*=0.16 | | | *P*<0.001 | | | *P*=0.419 | |

### 3.2 Evaluation of efficiency

Additionally, we performed an analysis of time for radiologists to label and annotate abnormalities with and without SIFT aids. With the help of SIFT, the efficiency can be improved by almost 50%. As shown in Table 4, the average review time without SIFT was 63.90s. The one with SIFT assistance was 36.85s, which reduced by 42.33%.

**Table 4.** Radiologists' annotation time with and without SIFT

| Review time (second) | Radiologist 1 | Radiologist 2 | Radiologist 3 | Average |
|---|---|---|---|---|
| Without SIFT assistance | 63.08 | 72.53 | 56.1 | 63.90 |
| With SIFT assistance | 30.48 | 47.17 | 32.9 | 36.85 |
| Improvement (%) | 51.68% | 34.96% | 41.35% | 42.33% |

### 3.3 Evaluation of Accuracy for Different Types of Abnormalities

In order to evaluate the accuracy of radiologists before and after the assistance, we conduct experiments on different types of abnormalities including fibrogenesis, pulmonary nodule, common pneumonia, secondary pulmonary tuberculosis and COVID-19. In the experiments, we define positive images as images containing specific abnormality and negative images as images containing other abnormality.

As shown in Table 5, radiologist's performance in labelling abnormalities statistically remain the same before and after use of SIFT. During real annotation, pathology of image is typically known except some comorbidity, radiologists are asked to label the center and annotate the boundary of abnormality. Radiologists will not be asked to label the abnormality. Therefore, radiologist's performance may not be the concern most of the time.

**Table 5**. AUC values of radiologists with and without SIFT aids in five different abnormalities

| | Without SIFT Aids (2 weeks ago) | | | With SIFT Aids | | | *P* value |
|---|---|---|---|---|---|---|---|
| Positive Case | Radiologist 1 | Radiologist 2 | Radiologist 3 | Radiologist 1 | Radiologist 2 | Radiologist 3 | |
| Pulmonary tuberculosis | 0.90 | 0.97 | 0.96 | 0.94 | 0.95 | 0.95 | 0.874 |
| Pulmonary nodule | 0.94 | 0.97 | 0.96 | 0.9 | 0.95 | 0.95 | 0.118 |
| Pneumonia | 0.86 | 0.90 | 0.92 | 0.87 | 0.91 | 0.88 | 0.728 |
| Covid-19 (GGO) | 0.68 | 0.73 | 0.69 | 0.73 | 0.73 | 0.69 | 0.423 |
| Fibrogenesis | 0.67 | 0.75 | 0.85 | 0.83 | 0.88 | 0.86 | 0.161 |
| **Average** | **0.82** | **0.87** | **0.88** | **0.86** | **0.89** | **0.87** | **0.37** |

SIFT is the system that will generate high-quality training, validation, and testing images for ML/AI. SIFT is not designed as computer aided detection/diagnosis/triage system used by clinicians. To summarize, radiologists' performance with and without SIFT are shown in Table 6. Consistency improves for annotated area and labelling. Without SIFT, radiologist annotated abnormality area is 29% larger than SIFT annotated area. Radiologists accept 93% SIFT annotated area and SIFT is able to reduce the inter-observer variation by 30%. Besides, consistency in labelling TP ROI and not labelling FP

ROI increases by 5.0% and 27.7% with SIFT Aids respectively. In the other aspect, efficiency increase by 42.33%. It only took SIFT 0.328 sec to process a single CXR on GPU or 3-5 sec on a CPU.

**Table 6.** Summary of radiologists' performance with and without SIFT

| | Measures (Average) | Without SIFT assistance | With SIFT assistance | Changes | *P* value |
|---|---|---|---|---|---|
| Consistency | Annotated area | 8,464,316 | 6,074,638 | -28.23% | *P* <0.001 |
| | Consensus area (IoU) | 0.629 | 0.762 | 25.27% | *P*<0.001 |
| | Consensus labelled abnormalities | | | | |
| | Number of consensus TP | 201 | 211 | 5.00% | *P*>0.001 |
| | Number of consensus FP | 65 | 47 | -27.73% | *P*<0.001 |
| Efficiency | Average time to annotate single image (s) | 63.90 | 36.85 | 42.33% | *P*<0.001 |
| Accuracy | Average AUC | 0.857 | 0.873 | 0.016 | *P*>0.001 |

From experiments, we find that potentially, the labelling and annotation time can be 10 times faster should rating (0 to 5) is not used to label the abnormality. More importantly, use of SIFT does not change the accuracy of labelling. The irregular shape and low contrast of abnormality will affect the efficiency and annotation consistency. Therefore, AI is expected to help radiologists more on irregular shape of object. SIFT has helped more for radiologist who does not have much of experience in labelling and annotating certain abnormality.

## 4. CONCLUSION

In this paper, we propose an AI-based Smart Imagery Framing and Truthing (SIFT) System which is designed to generate high-quality annotated cases to train machine learning and artificial intelligence. Evaluation of consistency, efficiency and accuracy for radiologists with and without SIFT was conducted. The results show that SIFT can significantly increase the radiologists' efficiency for the labeling and annotation of abnormalities. SIFT can also increase the radiologists' consistency by reducing the inter- observer's variation in labeling and annotating the abnormalities and areas. It reveals that our SIFT system can automatically annotate abnormalities to assist radiologists in generating high-quality image data labelling consistently and efficiently for the development of ML/DL in radiological applications.

## ACKNOWLEDGMENTS

This study has received funding by the National Key Research and Development Program of China (Grant No.: 2019YFE0121400), the Shenzhen Science and Technology Program (Grant No.: KQTD2017033110081833; JSGG20201102162802008; JCYJ20220531093817040), and the Shenzhen Fundamental Research Program (Grant No.: JCYJ20190813153413160). This research work was supported in part by the Lister Hill National Center for Biomedical Communications of the National Library of Medicine (NLM), National Institutes of Health.

## REFERENCES

[1] Chartrand G., Cheng P. M., Vorontsov E., Drozdzal M., Turcotte S., Pal C. J., Kadoury S., Tang A. Deep Learning: A Primer for Radiologists, Radiographics. 37(7), 2113-2131 (2017).

[2] Litjens G., Kooi T., Bejnordi B. E., Setio A. A. A., Ciompi F., Ghafoorian M., van der Laak J. A. W. M., van Ginneken B., Sánchez C. I. A survey on deep learning in medical image analysis, Med Image Anal. 42, 60-88 (2017).

[3] Zhou S. K., Greenspan H., Davatzikos C., Duncan J. S., Ginneken B. V., Madabhushi A., Prince J. L., Rueckert D., Summers R. M. A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises, proc. IEEE, 820-838 (2021).
[4] Rajpurkar P., Irvin J., Ball R. L., Zhu K., Yang B., Mehta H., Duan T., Ding D., Bagul A., Langlotz C. P., Patel B. N., Yeom K. W., Shpanskaya K., Blankenberg F. G., Seekins J., Amrhein T. J., Mong D. A., Halabi S. S., Zucker E. J., Ng A. Y., Lungren M. P. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists, PLoS Med. 15(11), e1002686 (2018).
[5] Wei Q., Ren Y., Hou R., Shi B., Lo J. Y., Carin L. Anomaly detection for medical images based on a one-class classification, proc. SPIE: Medical Imaging, 105751M (2018).
[6] Moradi M., Madani A., Karargyris A., Syeda-Mahmood T. Chest x-ray generation and data augmentation for cardiovascular abnormality classification, proc. SPIE: Medical Imaging, 57 (2018).
[7] Esmaeilzadeh S., Belivanis D. I., Pohl K. M., Adeli E. End-To-End Alzheimer's Disease Diagnosis and Biomarker Identification, Mach Learn Med Imaging. 11046, 337-345 (2018).
[8] Candemir S., Nguyen X. V., Prevedello L. M., Bigelow M. T., White R. D., Erdal B. S. Predicting rate of cognitive decline at baseline using a deep neural network with multidata analysis, J Med Imaging. 7(4), 044501 (2020).
[9] Wong K. C. L., Karargyris A., Syeda-Mahmood T., Moradi M. Building Disease Detection Algorithms with Very Small Numbers of Positive Samples, proc. Medical Image Computing and Computer Assisted Intervention − MICCAI 2017, 471-479 (2017).
[10] Tajbakhsh N., Shin J. Y., Gurudu S. R., Hurst R. T., Kendall C. B., Gotway M. B., Jianming L. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?, IEEE Trans Med Imaging. 35(5), 1299-1312 (2016).
[11] Esteva A., Robicquet A., Ramsundar B., Kuleshov V., DePristo M., Chou K., Cui C., Corrado G., Thrun S., Dean J. A guide to deep learning in healthcare, Nat Med. 25(1), 24-29 (2019).
[12] Park S. H., Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction, Radiology. 286(3), 800-809 (2018).
[13] Willemink M. J., Koszek W. A., Hardell C., Wu J., Fleischmann D., Harvey H., Folio L. R., Summers R. M., Rubin D. L., Lungren M. P. Preparing Medical Imaging Data for Machine Learning, Radiology. 295(1), 4-15 (2020).
[14] Yang F., Lu P. X., Deng M., Wáng Y. X. J., Rajaraman S., Xue Z., Folio L. R., Antani S. K., Jaeger S. Annotations of Lung Abnormalities in Shenzhen Chest X-ray Dataset for Computer-Aided Screening of Pulmonary Diseases, Data. 7(7), (2022).
[15] Abujudeh H. H., Boland G. W., Kaewlai R., Rabiner P., Halpern E. F., Gazelle G. S., Thrall J. H. Abdominal and pelvic computed tomography (CT) interpretation: discrepancy rates among experienced radiologists, Eur Radiol. 20(8), 1952-1957 (2010).
[16] Rozenberg E., Freedman D., Bronstein A. M. Localization with Limited Annotation for Chest X-rays, proc. Machine Learning for Health (ML4H) at NeurIPS, 1-9 (2019).
[17] Lin G., Kunlei H., Qian X., Lingjun Q., Stefan J., Bin Z., Shenwen Q., Li X., Guanxun C., Fleming Y.M.L. Developing and assessing an AI-based multi-task prediction system to assist radiologists detecting lung diseases in reading chest X-ray images, proc. SPIE: Medical Imaging, to be published (2023).