Nodule Generation in Chest Radiographs by Applying CT Templates

Marco Post
s1046670

Lennart Geertjes
s4606892

Reza Shokrzad
s1056369

***Abstract*—Deep learning is becoming more common in medical imaging systems. With the use of deep learning models, new challenges arise, such as the issue of data scarcity, for the selection of training data. In this research, we examine how an out-of-the-box detection algorithm can benefit from generated lung nodule data. A pipeline was created that produces realistic looking, simulated nodules and the accompanying metadata. The level of similarity to real nodules was tested with the use of detection algorithms. These tests show however that the generated nodules still need to be improved as they, in their current state, harm the performance of detection of real data.**

***Index Terms*—Deep learning, CT, Chest X-ray, data augmentation, simulation, nodules,**

I. Introduction

Lung cancer is one of the most common types of cancer for both genders and is the leading cause of cancer deaths worldwide [18]. Malignant nodules are a critical factor in this dangerous problem, when these nodules have enough time to reproduce before early-stage detection or patient complaints, the consequences can be fatal [14]. Various countries have opted for regular computerized tomography (CT) scans for cancer screening in high-risk individuals to catch these nodules before they can become metastasized. Even with CT scan screening, some research indicates that experts fail to find all visible nodules, with a sensitivity between 36% and 86% [9], [21], [24]. The idea of automated nodule detection has gained more attraction over the recent years for companies and academic groups [21]. Because of a recent increase in image processing techniques and the important role of these systems in diagnosis, there are diverse attempts to increase the accuracy of diagnosis. Sahiner et al. demonstrated that Computer-aided Diagnosis (CAD) had an enhancing effect on radiologist accuracy [12].

Even though CAD may have become more advanced over the last years, there are still numerous inherent challenges in finding appropriate automated algorithms for Chest radiograph analysis [17]. Such challenges are the number of required radiographs to train these CAD models, the complexity of a CAD model's interpretation and their proven significance in medical procedures. If these and other obstacles can be overcome, CAD models can become a low-constraint, high-impact tool with high levels of automation capable of aiding manual radiologists' work in the near future. Especially valuable in CAD model automation, is the ability to provide high sensitivity for subtle findings, less tedious daily tasks, and delivery of analysis in case of the absence of a radiologist. In this respect, deep learning is the most commonly applied approach in recent works [3]. Deep learning approaches allow for the modelling of complex problems, which perfectly suits a CAD scenario. As the detection of pulmonary nodules is a tremendously difficult task, even for expert radiologists. Therefore combining the complex nature of nodule detection and the data-hungry deep learning approach, makes having wide-ranging databases for this process a necessity [10]. However, the availability of these publicly available wide-ranging databases is limited at best. This is due to the time consuming and costly nature of labelling the radiographs by an expert hand, as well as the privacy-sensitive nature of its content. Furthermore, the nature of the problem causes a shortage of usable data, there are fewer cases of chest radiographs with pulmonary nodules than there are without. Making available data sets inherently unbalanced, causing issues in the training of deep learning CAD models as the models are prone to overfit on the clean radiographs. Consequently failing to properly predict the serious cases of lung cancer as they are underrepresented within the used training data. Attempts have been made to solve these issues by applying data augmentation [8], the commonly used data pre-processing approach. Data augmentation allows for the artificial tweaking of features within the data set, to force the model to pay more attention to those now augmented features. Applying this approach to detecting pulmonary nodules in radiographs using CAD, means enhancing the detection capability of the CAD model to detect nodules by augmenting nodules into clean radiographs. This approach could solve both the current shortage of nodule radiograph data and enhance existing CAD models ability to perform nodule detection tasks. This is the aim of the Grand ISMI Nodule Detection challenge. The Grand Challenge ISMI-NODE21 focuses on the detection and generation of nodules in chest radiographs.

This paper will focus on the generation track of the challenge. We hypothesize that the generation of artificial nodules, indistinguishable from real nodules, lies at the root of solving the data shortage issues within CAD model training. This hypothesis is reinforced by a growing research base, as displayed in the work of Zhao et al [25], in which

deep learning data augmentation methods are reported to increase performance by as much as 22.78%. The authors of this paper argue that augmenting data is one of the most important methods for performance-enhancing of current and future deep learning inspired CAD models. In this work, we propose a method to generate synthetic lesions and subsequently superimpose these generated nodules in radiographs by using CT scan templates of real nodules. This method is inspired by Litjens et al [10] and serves as a test run before the public release of the before mentioned Grand Challenge. We hypothesise that an already existing reliable deep learning nodule detection system will greatly benefit from a considerably larger training dataset, augmented with generated realistic-looking nodules. Both in terms of model performance, ease of training and testing. This in turn can facilitate early detection of malignant lung nodules, contributing to the prevention of cancer deaths worldwide.

II. Materials

To generate nodules, multiple sources of data were used [1], [2], [6], [15], [16], [22]. First of all 1186 CT patches containing nodules from the LIDC-IDRI dataset were used. These patches are 3 dimensional and have a size of 50x50x50 voxels. These CT patches were resampled such that every voxel is 1x1x1 mm. For each CT patch, there is a corresponding mask, which segmented the nodule from its context. Furthermore, there were 1031 frontal radiograph images available, that did not contain any nodules. For these radiographs, a mask was provided which segments the lungs from their context. All the radiographs were preprocessed to have a standardized size of 1024x1024 pixels. The radiographs were cropped around the lungs and have black padding if necessary. For each radiograph, the pixel spacing is available as well. Further there was a dataset available which contained 1134 chest radiographs that did contain nodules, and the metadata containing the locations of the bounding boxes.

III. Methods

To generate nodules on radiographs, we created a nodule generation pipeline in Python. The goal of the pipeline is to generate realistic-looking nodules that will enhance the performance of current nodule detection systems. The method used for this pipeline is a variation on the method from Litjens et al. [10] that includes four main steps named CT preprocessing, augmentation, lung mask preprocessing and superimposing. For this process, CT scans containing nodules and associated masks of the nodules are needed. Additionally, radiographs of lungs and associated masks of lungs are needed. The method can be described in the following way. Firstly, CT scans are scaled to a range between zero and one to make the voxel values positive which is needed for the superimposing step later on. Then, voxels of CT are segmented using the associated nodule mask. To enhance the speed of the following procedures the nodule is cropped from the segmented CT. The 3D cropped nodule is projected to a 2D image using a method proposed by Campo et al. [11]. Afterwards, the 2D projected nodule is resampled using interpolation technique to an isotropic resolution equal to the radiograph. As augmentation, the projected nodule is either rotated 90 degrees flipped vertically or horizontally or zoomed in. For lung mask preprocessing the lung mask is cropped based on the diameter of the resampled projected nodule. A random location in the cropped lung is chosen to superimpose the nodule on after which the nodule can be projected resulting in a radiograph containing a nodule. The radiograph is then scaled to a 0 to 255 range as the original dataset we supplemented with this generated data was also in this range. To test the results we trained a detection model on real data and generated data, and compared outcomes.

A. Preprocessing

The first step of the pipeline is the preprocessing of the nodules. Preprocessing consists of five distinctive sequential steps: CT segmentation, cropping the part of CT including nodule, cropped CT projecting, 2D projected CT resampling and normalizing the 2D nodule. Figure 1 depicts the preprocessing steps visually. Further explanation for each mentioned step is provided in subsections of the method.

1) CT Segmentation: To segment nodules in a CT, we imposed a condition leaning on the state of the mask in that CT. If each pixel is a mask, it would replace with the content of CT. Otherwise, it would substitute with the background.

2) Cropping the CT containing 3D nodule: Before projecting the nodule to the 2D version, cropping the nodules in the CT is required. As a nodule is a small part of a patch, most of a CT would be the background (Figure 2(a)) that affects the brightness of imposed nodule in the following step (influence on the mean of pixel values). Moreover, cropping is an essential step in the prepossessing procedure for removal of the unworkable borders (figure 2(b)). Cropping also speeds up the remaining processes in the sequence as it decreases data to be processed.

3) Projecting the Cropped CT: For the projection of the cropped segmented 3D image of the nodule to a digitally reconstructed 2D image, the following transformation of Campo et al. [11] is used

$$X_P\left(x_i, x_j\right) = \frac{1}{N}\sum_{k=1}^{N} e^{\beta \frac{\max(X_{CT}(x_i, y_j, z_k), -1024)+1024}{1000}} \tag{1}$$

where X_{CT} is the 3D matrix describing the cropped CT and X_P is this matrix projected to 2D which describes the 2D image of the nodule. A $\beta = 0.85$ was chosen, as described in Campo et al. [11]

4) 2D Nodule CT Resampling: The goal of the resampling stage is to scale the CT patches to an adjustable size of a CXR. The process is based on a double transition from CT scale to an intermediate-range related CXR and converting them lastly to CT patches.

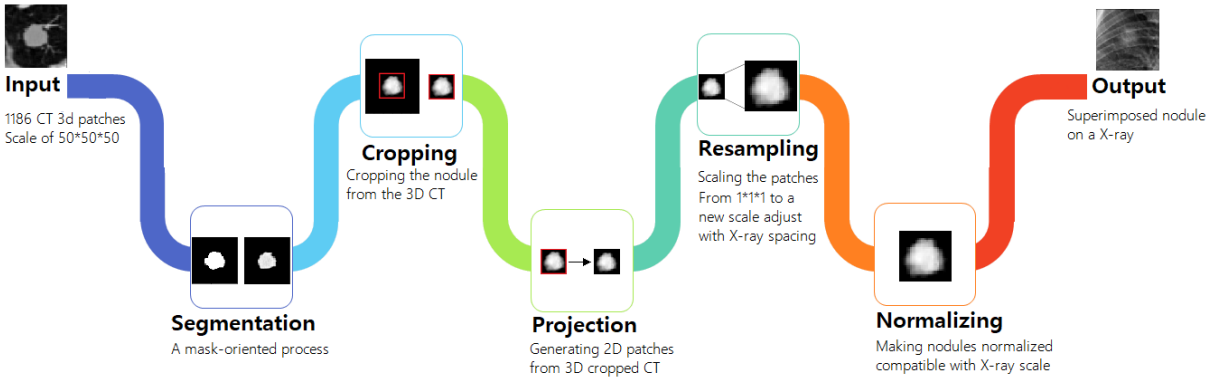

Fig. 1: Preprocessing Phase of Generating Synthetic Nodule

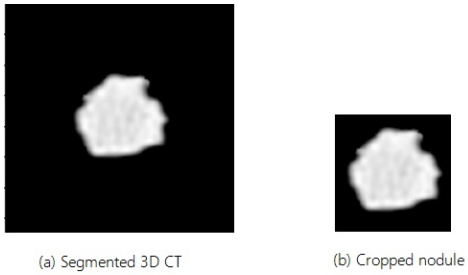

Fig. 2: Cropping process to remove useless borders

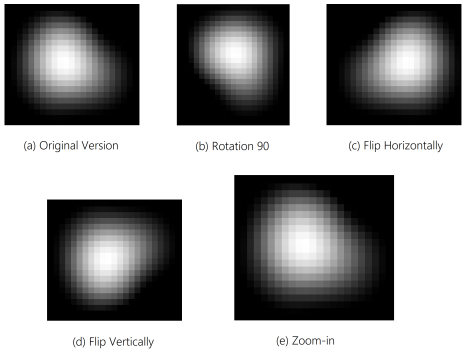

Fig. 3: Augmented Data Views

5) Normalizing the Nodule: The 2D nodule is scaled to the range of the values present in the x-ray on which it will be superimposed. This is necessary for the superimposition algorithm that is used.

B. Data Augmentation

To enhance the number of generated data a data augmentation strategy was applied in four different patterns. Since new synthetic data needs to be practical, the options of augmentation methods are limited. Rotation for 90°, vertical and horizontal flipping, and zooming-in was these four methods are shown in figure 3. To be more sensible, we weighted twice to original data. Therefore, the odds of augmented cases appearing are almost 17% and this number for the original version is 35%.

C. Lung mask preprocessing

While generating the radiographs, a location must be chosen to superimpose the 2D nodules. To accomplish this a random location within the lung mask is chosen. A random location must be chosen within the corresponding radiograph mask. To prevent random locations to be chosen where the superimposed nodule would fall outside of the lungs, namely at the edges, the lung mask is cropped as can be seen in figure 4. The yellow part of the lung mask is the area where the nodule can be placed. The green part is cropped from the mask. The thickness of it is half the diameter of the nodule.

Fig. 4: Preprocessed lung mask.

D. Superimposition

Inserting the resampled projected pattern into a converted radiograph is the last step in the generation pipeline. To inserting nodules in radiographs we applied the Perfect Spherical Nodule Simulation Method [13]. According to this method, a nodule contrast representation would be based on an empirical mathematical formula (equation 2).

$$c[r] = C(\frac{4}{D^4}r^4 - \frac{4.2}{D^2}r^2 + 1) \quad (2)$$

To be more specific, r stands for the radius that shows the distance of each pixel to the centre of the nodule. Parameters

C and D are the peak of the contrast value and the diameter of the nodule, respectively. In this term, C is a coefficient to regulate the blending process. There are several formulas in the literature to calculate the peak of contrast (C) however, according to our results, we considered it a free parameter and through testing came up with the value of 0.25. The output of this formula is a value for each pixel of the nodule that multiplies in its corresponding pixel value. According to this method, the intensity of brightness is based on the radius of the pixel that makes the superimposing process more realistic looking by seamlessly blending the superimposed nodule in its surroundings.

E. Detection

To test the performance-enhancing capabilities of our generated nodules, we tested the data on an already existing deep learning detection method. For the detection model, we chose Faster r-cnn [26]. We used an existing implementation of the model from Pytorch, so we would not be forced to build a detection system from the ground up. We also used an existing training pipeline from as a basis.

Before training, the 1134 radiographs of real data were split up into a training set, a validation set and a test set following a 90-5-5 split. This resulted in a training set of 930 images, a validation set of 52 images and a test set of 52 images. We then generated 930 radiographs with fake nodules for a second training. This meant that both training sets were the same size which would make it easier to compare results.

The model was trained multiple times on different types of data. First, the model was trained on only the real data, the outcome of this would serve as a baseline performance. Second, a model was trained on only generated data. Lastly, the model was trained on a dataset that was 50 per cent real data and 50 per cent generated data. An important note on this training set is that it is twice the size of the other two data sets and thus it has more update steps for each epoch.

The parameters stayed the same across all three training sessions. we used a learning rate of 0.005, a momentum of 0.9 and a weight decay of 0.0005. The models were trained for 10 epochs. Afterwards, the models are evaluated on the test set, over which the mean average precision (mAP) will be calculated. The range of Intersection over Union (IoU) values used for the mAP is 0.5 to 0.95. All boxes with a confidence score higher than 0.05 are analyzed. For the evaluation, we used code from the CocoAPI [27].

IV. Results

In this part, two distinctive outputs are reported. These are the final output of inserted nodules in radiographs and the result of training the neural network. Figure 5 depicts some examples of the superimposed nodules in radiographs.

The mAP of the three models we trained, tested on 52 radiographs containing only real data, is as follows:

Something noticeable during the training phase was that when training the model with generated data, the validation score increased after each epoch while the training score decreased initially and then stabilized.

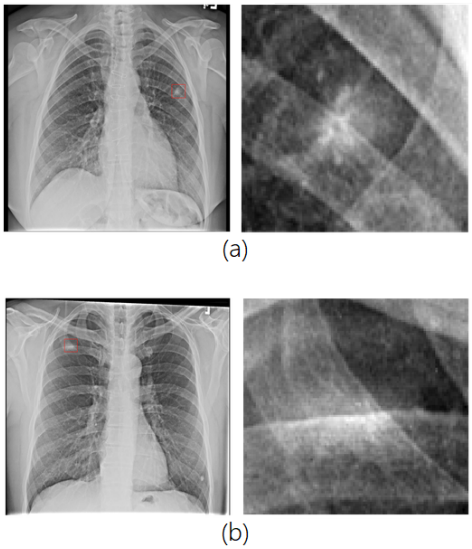

Fig. 5: Examples of superimposed nodules. The red square points out where the nodule is located

Dataset	mAP
Real	0.101
Generated	0.0
Mixed	0.039

V. Discussion

Detecting early-stage malignant lung cancer nodules is key for the prevention of numerous cancer deaths worldwide. Deep learning methods facilitate detection and investigation with great precision, leading to fast and efficient decision making in response to malignant nodules in the lungs. The increasing role of deep learning models for detection requires huge amounts of data, that can not yet be provided from natural sources. Data generation augmentation techniques are introduced to deal with the problem of data scarcity in medical imaging, and to potentially achieve an unprecedented level of robustness in modern nodule detection deep learning algorithms. The purpose of this investigation was to develop a pipeline for the generation of additional radiographs containing nodules, to ultimately enhance the performance of deep learning nodule detection systems. A pipeline was established by investigating and implementing the most applicable tools from the Litjens et al. [10] paper. The parameters of the tools and methods were examined empirically in more detail.

We believe that we made a good effort to approach realistic nodules. All steps in the pipeline of Litjens et al. were implemented to a certain degree of success. By adjusting the order of steps we were also able to speed up the process and we were also able to implement some simple forms of data augmentation. Further we made sure that all nodules would be super imposed completely within the lungs. However, to make claims regarding the 'reallness' of our generated nodules we implemented a detection model.

We chose to go for an existing Pytorch implementation of Faster r-cnn, which we trained on both real and generated

data. Unfortunately, the results show that when training the model using only generated data or expanding a dataset of real data with generated data, decreases the performance of the model when tested on real data. This means that our data, as it is now, is not yet suitable to train detection models with. This means that our current pipeline should be tweaked more, and more rigorous testing has to be performed before we can produce nodules that can be used to compliment real data sets. As it stands now, there exist subtle differences between real nodules and generated nodules, unfortunately this means that a system could overfit on these differences which harms its ability to detect real nodules. This is also what we saw during training. The more the system improved on detecting the generated nodules(measured through training loss), the worse it performed on real data (measured through validation loss).

There are also remarks to make concerning the results of the detection model. First of all the model did not seem to converge well during the training phase. Even when only using real data. It would have been more beneficial if the model had been optimized more for the task, as this would indicate the potential of the system and how far off a model trained on (a partially) generated dataset. We did some testing with different learning rates, and learning rate schedulers, however, the system stopped at a limited performance each attempt. Most of the time it stopped improving on the validation and training set after 4-6 epochs. Which is very early. More research needs to be done on optimizing the training of the model.

Furthermore, after making improvements such that our detection system would provide more meaningful feedback, it would be useful to look into generating a free-response ROC curve. Such a curve is very useful in detection tasks. Besides this, it is easier interpretable for clinical staff as it plots the sensitivity of a system against the average number of false positives per scan, which is more intuitive than the mAP score.

Given better availability of resources, we would also experiment with different generated nodules, for example, make specific nodules that are larger, smaller, spherical or even speculated. In real-life medical situations, all of the above-mentioned nodule variations would be possible, so it would be best to train a given CAD system on all the possible real-life circumstances, making the system as robust and practicable as possible. In addition, we would like to experiment more with the location of the simulated nodule. In this work we chose a random point in the lung mask, however, we aim to move it to a correlated point within the lungs, based on positions found, studying nodule positions in future research, to increase the ”realness” of the nodule. Besides, current nodule detection systems and even expert radiologists face the added challenge of detecting nodules located behind the heart or diaphragm. When generating nodules in future work, we can deal with this problem beforehand, making an effort to simulate nodules in these challenging areas, which can result in more detection focus and robustness of a model in problematic areas of the lungs.

Our original plan and literature review for this project was based on generative adversarial networks (GANs). We believe that it is entirely possible and feasible to achieve realistic looking simulated nodules with a GAN approach. We found it a promising approach to the problem, as described in various works [4], [5], [7]. However, due to time and resources constraints for this project and the demand for results, we opted to go for a variation of the method described in the paper of Litjens et al. instead of using GANs. However, we argue that the GAN approach to the problem of nodule simulation is a promising research direction for future investigation. The current investigation provides good information about how simulated nodules, aside from real data can be used in enhancing the performance of detection systems. Which in turn could prove useful to, especially, the preprocessing steps, but ultimately to the simulation to detection process as a whole when using GANs.

Ideally, we would like millions of nodules to train deep learning systems on, and we can only imagine what this would mean for the future of fake nodules detection training datasets. The most appealing future solution for this problem is of course concentrating more efforts on preventing diseases altogether, especially for lung cancer, tempering the ambitions of tobacco companies, even more, would result in an overall more healthy population [23]. However, we speculate that these societal changes take considerably longer than the potential changes possible in the healthcare systems around the world and the Netherlands. One of the biggest limitations, aside from trust in CAD systems by consumers, is the current power of the used hardware [20]. Although lately Moore’s Law is questioned, it is highly related to this issue. Stating that computer power will double every two years, this might not be every two years anymore, but it’s still improving [19]. This improvement will allow us to increase the capabilities of deep learning methods without memory issues and infeasible computation times. Thus allowing us to train with even more data at the same time. More data will allow us to learn even better models that are greatly appropriate for nodule detection systems. We hypothesize that with small incremental improvements, deep learning models will in the far future obtain a new position, where CAD systems are the golden standard for a trusted diagnosis in modern healthcare systems.

REFERENCES

[1] Andrey Fedorov, Matthew Hancock, David Clunie, Mathias Brochhausen, Jonathan Bona, Justin Kirby, John Freymann, Steve Pieper, Hugo Aerts, Ron Kikinis1, Fred Prior, 2019. Standardized representation of the LIDC annotations using DICOM. The Cancer Imaging Archive. doi: 10.7937/TCIA.2018.H7UMFURQ

[2] Bustos, A., Pertusa, A., Salinas, J.M., de la Iglesia-Vaya, M., 2020. PadChest: ´ A large chest x-ray image dataset with multi-label annotated reports. Medical Image Analysis 66, 101797. doi:10.1016/j.media.2020.101797.

[3] Çallı, E., Sogancioglu, E., van Ginneken, B., van Leeuwen, K. G., & Murphy, K. (2021). Deep Learning for Chest X-ray Analysis: A Survey. Medical Image Analysis, 102125.

[4] Changhee Han, Yoshiro Kitamura, Akira Kudo, Akimichi Ichinose, LeonardoRundo, Yujiro Furukawa, Kazuki Umemoto, Yuanzhong Li, and Hideki Nakayama. Synthesizing diverse lung nodules wherever massively: 3d multi-conditional gan-based CT image augmentation for object detection, 2019.

[5] Dakai Jin, Ziyue Xu, Youbao Tang, Adam P. Harrison, and Daniel J. Mol-lura. Ct-realistic lung nodule simulation from 3d conditional generative adversarial networks for robust lung segmentation, 2018.

[6] Demner-Fushman, D., Antani, S., Simpson, M., Thoma, G.R., 2012. Design and Development of a Multimodal Biomedical Information Retrieval System. Journal of Computing Science and Engineering 6, 168–177. doi:10.5626/JCSE.2012.6.2.168.

[7] Eric Wu, Kevin Wu, David Cox, and William Lotter. Conditional infilling gans for data augmentation in mammogram classification, 2018

[8] Hussain, Z., Gimenez, F., Yi, D., & Rubin, D. (2017). Differential data augmentation techniques for medical imaging classification tasks. In AMIA Annual Symposium Proceedings (Vol. 2017, p. 979). American Medical Informatics Association.

[9] Li, F., Arimura, H., Suzuki, K., Shiraishi, J., Li, Q., Abe, H., ... & Doi, K. (2005). Computer-aided detection of peripheral lung cancers missed at CT: ROC analyses without and with localization. Radiology, 237(2), 684-690.

[10] Litjens, G.J.S., et al. (2010) 'Simulation of Nodules and Diffuse Infiltrates in Chest Radiographs Using CT Templates'

[11] Monica I.C., Pascau J., Estepar R.S.J. (2018) 'Emphysema quantification on simulated X-rays through deep learning technique'

[12] Sahiner, B. et al. (2007) 'Malignant and benign breast masses on 3D US volumetric images: Effect of computer-aided diagnosis on radiologist accuracy', Radiology, 242(3), pp. 716–724.

[13] Samei, E., Flynn, M. J. and Eyler, W. R. (1999) 'Detection of subtle lung nodules: Relative influence of quantum and anatomic noise on chest radiographs', Radiology, 213(3), pp. 727–734.

[14] Saraswathi, S., Sheela, L. (2017). Detection of Juxtapleural Nodules in Lung Cancer Cases Using an Optimal Critical Point Selection Algorithm. Asian Pacific journal of cancer prevention: APJCP, 18(11), 3143–3148.

[15] Setio et al., Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images:: The LUNA16 challenge, Medical Image Analysis 42, DOI:: 10.1016/j.media.2017.06.015

[16] Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K.i., Matsui, M., Fujita, H., Kodera, Y., Doi, K., 2000. 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. American Journal of Roentgenology 174, 71–74. doi:10.2214/ajr.174.1.1740071.

[17] Sogancioglu, E. et al. (2021) 'Deep Learning for Chest X-ray Analysis: A Survey'.

[18] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 71(3), 209-249.

[19] "A New Era of Innovation: Moore's Law Is Not Dead and AI Is Ready to Explode." SiliconANGLE, 17 Apr. 2021, siliconangle.com/2021/04/10/new-era-innovation-moores-law-not-dead-ai-ready-explode/.

[20] Sze, V., Chen, Y. H., Emer, J., Suleiman, A., & Zhang, Z. (2017, April). Hardware for machine learning: Challenges and opportunities. In 2017 IEEE Custom Integrated Circuits Conference (CICC) (pp. 1-8). IEEE.

[21] Toyoda, Y. et al. (2008) 'Sensitivity and specificity of lung cancer screening using chest low-dose computed tomography', British Journal of Cancer, 98(10), pp. 1602–1607.

[22] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M., 2017b. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in IEEE Conference on Computer Vision and Pattern Recognition, pp. 2097–2106. doi:10.1109/cvpr.2017.369.

[23] World Health Organization. (2020). Tobacco.

[24] Quekel, L. G., Kessels, A. G., Goei, R., van Engelshoven, J. M. (2001). Detection of lung cancer on the chest radiograph: a study on observer performance. European journal of radiology, 39(2), 111-116.

[25] Zhao, D., Zhu, D., Lu, J., Luo, Y., Zhang, G. (2018). Synthetic medical images using F&BGAN for improved lung nodules classification by multi-scale VGG16. symmetry, 10(10), 519.

[26] Ren, S. He, K., Girshick, R., Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39, 1137-1149.

[27] Lin, T., Maire, M., Belongie, S, Bourdev L., Girshick, R., Hays J., Perona, P., Ramanan, D., Zitnick C.L., Dollár, P. (2014) Microsoft COCO: Common Objects in Context, Computer Vision – ECCV 2014, 740-755.