2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)
April 3-7, 2020, Iowa City, Iowa, USA

# MITIGATING ADVERSARIAL ATTACKS ON MEDICAL IMAGE UNDERSTANDING SYSTEMS

*Rahul Paul*[1], *Matthew Schabath*[2], *Robert Gillies*[3], *Lawrence Hall*[1], and *Dmitry Goldgof* [1]

[1]Computer Science and Engineering Department, University of South Florida, Tampa, FL
[2]Department of Cancer Epidemiology, H. L. Moffitt Cancer Center and Research Institute, Tampa, FL, USA
[3]Department of Cancer Physiology, H. L. Moffitt Cancer Center and Research Institute, Tampa, FL, USA

*Abstract*—**Deep learning systems are now being widely used to analyze lung cancer. However, recent work has shown a deep learning system can be easily fooled by intentionally adding some noise in the image. This is called as Adversarial attack. This paper presents an adversarial attack for malignancy prediction of lung nodules. We found that the adversarial attack can cause significant changes in lung nodule malignancy prediction accuracy. An ensemble-based defense strategy was developed to reduce the effect of an adversarial attack. A multi-initialization based CNN ensemble was utilized. We also explored adding adversarial images in the training set, which eventually reduced the rate of mis-classification and made the CNN models more robust to an adversarial attack. A subset of cases from the National Lung Screening Trial (NLST) dataset were used in our study. Initially, 75.1%, 75.5% and 76% classification accuracy were obtained from the three CNNs on original images (without an adversarial attack). Fast Gradient Sign Method (FGSM) and one-pixel attacks were analyzed. After the FGSM attack, 46.4%, 39.24%, and 39.71% accuracy was obtained from the 3 CNNs. Whereas, after a one pixel attack 72.15%, 73%, and 73% classification accuracy was achieved. FGSM caused much more damaged to CNN prediction. With a multi-initialization based ensemble and including adversarial images in the training set, 82.27% and 81.43% classification accuracy were attained after FGSM and one-pixel attacks respectively.**

## 1 Introduction

Lung cancer is the world's most frequently diagnosed cancer with a 5-year survival of 18% [1]. To enhance the survival and patient outcome, early diagnosis of lung cancer is a priority. Computed tomography (CT) scan is the most standard imaging approach for lung cancer screening.

The use of Convolutional Neural Network (CNN) has grown enormously in recent years and has also recently been explored for cancer diagnosis. The deep learned CNN may be fooled by a small amount of intentional distortion of pixel value(s). These distorted images are referred to as adversarial images [2]. The machine learning models trained and tested on an undistorted dataset may misclassify adversarial images. These distorted images can also deceive human recognition [3]. Intentional pixel distortion can also be applied to CT scans of lung nodules and may also impact the malignancy prediction.

Adversarial attacks on lung nodule CT images were analyzed in this study. We made the following contributions,

1. Analyzed two adversarial attacks on lung nodules CTs.

2. We also proposed a simple defense approach by incorporating adversarial examples in the training set to mitigate the effect of adversarial attack.

3. A multi-initialization CNN ensemble was proposed to enhance the malignancy prediction. Current results were compared with our previous study on lung nodule malignancy prediction [4].

## 2 Adversarial Attacks and defense strategies

Szegedy et al. [3] first revealed that the state-of-the-art CNN can be fooled with small perturbation in images. These perturbed images are called adversarial images. A recent study by Mirsky [5] showed how a person could alter a CT scan by infiltrating the picture archiving and communication system (PACS). Finlayson [6] analyzed adversarial attacks on medical images using Projected Gradient Descent (PGD) and a patch-based attack. These studies focused on adversarial attacks.

In our current study, we investigated untargeted adversarial attacks that could be performed after generation of scans and won't need to add or remove any portion of a tumor on the CT scans. The changes would be indiscernible to a physicians'eyes, but can fool a CNN. We utilized the FGSM and one-pixel attack on the lung nodules. We also explored novel defense strategies against adversarial attacks.

### 2.1 Fast Gradient Signed Method

The fast gradient signed method (FGSM) is one of the first and influential adversarial attacks proposed by Goodfellow et al. [7]. The FGSM utilizes the gradient of the loss of an input image to generate an adversarial image maximizing the loss. The FGSM approach can be summarized as,

$$advI = I + \epsilon * sign(\nabla_I * J(\theta, I, Y)) \tag{1}$$

where $advI$ is the adversarial image, $I$ is the original Image, $Y$ is the image label, $\epsilon$ is the amount of perturbation to be added, $\theta$ is the CNN parameters and $J$ is the loss function.

Our intention was to analyze if the model prediction could also be affected after adding very little noise. Thats why we chose an epsilon of 0.01 [8].

### 2.2 One-Pixel Attack

Su et al. [9] experimented with altering different numbers of pixels and showed that with a change of just one pixel, the CNN prediction could be altered. The differential Evolution (DE) algorithm is an evolutionary algorithm, which was used to chose one pixel iteratively. At first, by modifying a random pixel, several adversarial images are generated and then a CNN model is used to check the prediction. Then by combining the position of the pixels from the previous stage and colors, more adversarial images are generated and again a CNN model is used to analyze the prediction. Now for any pixel that

978-1-5386-9330-8/20/$31.00 ©2020 IEEE

lowered the CNNs'prediction from the first step, replace those pixels with the current value and repeat for a few iterations. After the final iteration, the pixel value that most changed the CNNs'prediction will be returned as the attack image [10].

### 2.3 *Defense Against Adversarial Attacks*

Researchers have proposed various defense strategies [11]–[13] to boost the robustness of a CNN model against adversarial attack. Goodfellow et al. [7] added adversarial examples in the training set to improve the classification performance for the MNIST dataset. Tramer et al. [14] proposed an ensemble adversarial training approach with a single classifier, not a traditional multiple classifier predictor. They incorporated adversarial samples for training from different CNNto enhance the diversity and robustness.

We proposed a defense strategy by combining adversarial examples in the training set. We further enhanced the classification performance by creating an ensemble of CNNs.

## 3 Convolutional Neural Network

A Convolutional Neural Network (CNN) [15] is a variant of neural network which is using widely for image classification and recognition. We designed three CNN architectures using Keras [16] and Tensorflow [17]. RMSprop optimizer [18] was chosen to train the CNNs with a learning rate of 0.0001 and a batch size of 16. Each of the CNNs was trained for 200 epochs. Overfitting is very common for a CNN. To reduce overfitting dropout [19] along with L2 normalization [20] was applied before the final classification layer. In the final classification layer, the sigmoid activation was used as we have 2 classes (incident lung cancer and control cases). The parameters of the CNN architectures are shown in Table 1. Detailed analysis of the CNN architectures and parameters can be found in [4].

## 4 NLST Dataset

National Lung Screening Trial (NLST) was a multi-institutional study spanning three years. In the first year, there was a baseline scan (T0) and two follow-up scans (T1 and T2) in the next two years with a one-year gap. From the National Cancer Institutes (NCI) cancer data access system, a de-identified subset of malignant and control cases [21] from the CT arm of the NLST [22] was selected for our study. We conducted our analysis with the CT scans from the baseline scan (T0). These cases from the baseline scan was divided into two Cohorts: Cohort 1 (training set) and Cohort 2 (test set) and each cohort had two classes: incident lung cancer and control cases. Cohort 1 included cases which had a positively screened nodule during the baseline scan (T0) and after the first follow-up scan (T1) some of these positively screened nodules were found to be malignant. Cohort 1 had 85 incident lung cancer and 176 control cases. Cohort 2 included cases where some of the positively screened nodules from the baseline scan (T0) were found to be malignant after the second follow-up scan (T2), i.e. approximately two years after the baseline scan. Cohort 2 had 85 incident lung cancer and 152 control cases. Details about the selection of Cohorts can be found in [23].

Definiens Software [24] was used by a radiologist with more than 9 years of experience to segment the nodules from the CT scans [23]. For the two-dimensional CNN (CNN applied on 2-D images), we chose a slice with the largest nodule area for each patient and only the nodule region was extracted.

## 5 Experiments and Results

Cohort 1 and Cohort 2 were chosen as the training set and test set, respectively. Cohort 1 randomly was divided into training (70%) and validation set (remaining 30%). After that rotation of 12-degrees and then vertical flipping were applied on both training and validation. Cohort 2 was kept completely separate for final model evaluation using accuracy and area under the receiver-operator curve (AUCROC) [25].

### 5.1 *Adversarial Attack on Original Lung Nodule Images*

We achieved 75.1% (0.84 AUC), 75.5% (0.86 AUC), and 76% (0.87 AUC) accuracy using CNN architecture 1, 2 and 3, respectively without adversarial attack. Results obtained after the adversarial attack were compared with the original results.

FGSM (epsilon = 0.01) and a one-pixel attack was carried out on the lung nodule images. CNN architecture 1 was used to generate the attack images. Examples of adversarial attack images are shown in Figure 1.

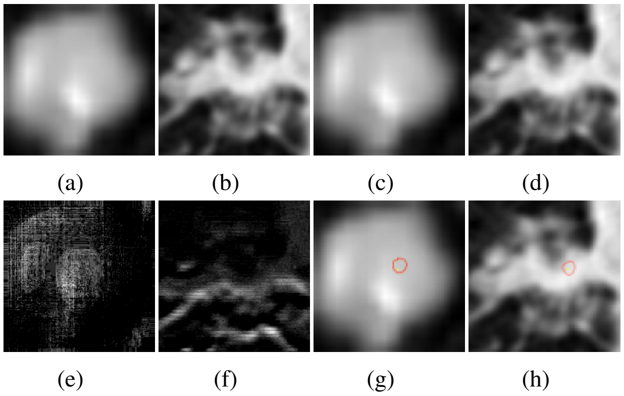

**Fig. 1**: (a,b) Original nodules; (c,d) FGSM images; (e,f) Difference of original and FGSM images; (g,h) 1-pixel attack images (the pixel value changed is shown in red)

A CNN model trained on Cohort 1 (no adversarial attack was performed on Cohort 1) was tested on an unseen Cohort 2 (adversarial attack was performed here).

FGSM attack images were generated only for Cohort 2. We obtained 46.4% (0.38 AUC), 39.24% (0.36 AUC), and 39.71% (0.42 AUC) classification accuracy using CNN architecture 1, 2 and 3, respectively. This resulted in a 28%, 36%, and 36% decrease in classification accuracy using CNN architecture 1, 2 and 3, respectively. Using the DE algorithm, one-pixel attack images for only Cohort 2 were obtained. For the DE algorithm, five iterations were used. We observed 72.15% (0.7 AUC), 73% (0.69 AUC) and 73% (0.72 AUC) classification accuracy using CNN architecture 1, 2 and 3, respectively. For one-pixel attack, the accuracy didn't drop much, which indicates that our CNNs were less affected by one pixel change.

**Table 1**: CNN architectures and parameters

| CNN architecture 1 | | CNN architecture 2 | | CNN architecture 3 | |
|---|---|---|---|---|---|
| Layers | Parameters | Layers | Parameters | Layers | Parameters |
| Input | 100 x100 | Input | 100 x100 | **Left BRANCH** | |
| Conv1 | 64x5x5,pad 0,stride 1 | Conv1 | 64x5x5,pad 0,stride 1 | Input | 100 x100 |
| Leaky ReLU | alpha=0.01 | Leaky ReLU | alpha=0.01 | Max Pool 1 | 10x10 |
| Max Pool 1 | 3x3, stride 3, pad 0 | Max Pool 1 | 3x3, stride 3, pad 0 | Dropout | 0.1 |
| Conv2 | 64x2x2,pad 0,stride 1 | Conv2 | 64x2x2,pad 0,stride 1 | **Left BRANCH** | |
| Leaky ReLU | alpha=0.01 | Leaky ReLU | alpha=0.01 | Conv1 | 64x5x5,pad 0,stride 1 |
| Max Pool 2 | 3x3, stride 3, pad 0 | Max Pool 2 | 3x3, stride 3, pad 0 | Leaky ReLU | alpha=0.01 |
| Dropout | 0.1 | Dropout | 0.1 | Max Pool 2 | 3x3, stride 3, pad 0 |
| FC 1+ ReLU | 128 | FC 1+ ReLU | 128 | Conv2 | 64x2x2,pad 0,stride 1 |
| FC 2+ ReLU | 8 | LSTM 1 + ReLU | 8 | Leaky ReLU | alpha=0.01 |
| L2 regularizer | 0.01 | L2 regularizer | 0.01 | Max Pool 3 | 3x3, stride 3, pad 0 |
| Dropout | 0.25 | Dropout | 0.25 | **Merge Left and Right** | |
| FC 3+ Sigmoid | 1 | FC 3 + Sigmoid | 1 | Conv 3 | 64x2x2, pad 0, stride 1 |
| | | | | Max Pool 4 | 2x2, stride 2, pad 0 |
| | | | | L2 regularizer | 0.01 |
| | | | | Dropout | 0.1 |
| | | | | FC 1+ Sigmoid | 1 |
| Total parameters | 841,681 | Total parameters | 845,033 | Total parameters | 39,553 |

**Table 2**: Classification performance of adversarial attacks using three CNNs trained using seven different seed point initialization: Train on Cohort 1 Original images only and Test on adversarial Cohort 2 images

| CNN 1 | | | CNN 2 | | | CNN 3 | | |
|---|---|---|---|---|---|---|---|---|
| Original (no attack) | FGSM | 1-pixel attack | Original (no attack) | FGSM | 1-pixel attack | Original (no attack) | FGSM | 1-pixel attack |
| 75.1% | 46.4% | 72.15% | 75.5% | 39.2% | 73% | 76% | 39.71% | 73% |
| 73.4% | 49.4% | 71.3% | 73.83% | 41.77% | 70.88% | 75.1% | 40.56% | 71.3% |
| 74.26% | 49.4% | 71.72% | 74.68% | 54% | 69.62% | 76% | 37.15% | 73.4% |
| 74.26% | 51.5% | 71.3% | 74.26% | 39.2% | 72.5% | 74.26% | 37.15% | 70.04% |
| 75.1% | 52.74% | 70.04% | 74.68% | 49.4% | 71.3% | 75.5% | 39.71% | 72.5% |
| 73.83% | 48.1% | 71.3% | 75.1% | 54.33% | 72.5% | 73.83% | 36.3% | 72.15% |
| 73% | 39.24% | 72.5% | 74.68% | 43.24% | 72.15% | 75.5% | 40.56% | 72.5% |

**Table 3**: Ensemble of CNNs: Train on Cohort 1 Original images only and Test on adversarial Cohort 2 images

| Approach | CNN 1: Ensemble of 7 CNNs | CNN 2 : Ensemble of 7 CNNs | CNN 3: Ensemble of 7 CNNs | Overall : Ensemble of 21 CNNs |
|---|---|---|---|---|
| Original | 84.8% (0.89 AUC) | 87.34% (0.89 AUC) | 87.34%(0.9 AUC) | 90.29% (0.94 AUC) |
| FGSM | 62.86% (0.41 AUC) | 62.44% (0.44 AUC) | 66.67% (0.53 AUC) | 67.5% (0.5 AUC) |
| 1-pixel attack | 76.3% (0.77 AUC) | 77.21% (0.77 AUC) | 77.21% (0.82 AUC) | 78.9% (0.84 AUC) |

**Table 4**: Train Cohort 1 using Original and adversarial images and test on Original Cohort 2 images

| CNN 1 | | CNN 2 | | CNN 3 | |
|---|---|---|---|---|---|
| FGSM | 1-pixel attack | FGSM | 1-pixel attack | FGSM | 1-pixel attack |
| 73% | 73% | 74.68% | 73.83% | 74.26% | 73.83% |
| 73.4% | 72.5% | 74.68% | 72.15% | 73.83% | 72.5% |
| 73.83% | 73% | 74.26% | 72.5% | 73% | 74.26% |
| 73.83% | 71.3% | 74.68% | 73.4% | 73.4% | 71.72% |
| 73.4% | 70.04% | 74.26% | 72.15% | 74.26% | 73.83% |
| 72.5% | 72.5% | 71.72% | 73% | 74.26% | 73.4% |
| 72.5% | 73% | 71.72% | 72.5% | 73.4% | 74.26% |

### *5.2 Defending Against Adversarial Attacks*

We trained each CNN architectures multiple times (seven) with a different seed point for weight initialization to verify the adversarial attack's stability and the performance variability. An ensemble defense strategy was applied using multiple CNNs to enhance the classification performance and robustness of individual CNN architecture. For comparison we analyzed the performance of CNNs on original Cohort 2 images. Detailed results obtained from 3 CNNs after training using seven initializations are shown in Table 2. Then, we created an ensemble by averaging the pseudo probability output from 21 CNNs (3 CNNs with 7 initialization). An improvement against the adversarial attacks was observed. Using the ensemble, 67.51% classification accuracy were obtained which was a 12-25% improvement from the results shown in Table 3.

Then, the stability of the adversarial attack and the variability of CNN's classification was analyzed for one-pixel attack. Detailed results obtained from 3 CNNs after training using seven initializations are shown in Table 2. Then, we created an ensemble by averaging the pseudo probability output from 21 CNNs. 78.9% accuracy was achieved using the ensemble, which was an improvement on the results shown in Table 2. The ensemble results are shown in Table 3.

In our study we generated adversarial images (both FGSM

**Table 5**: Ensemble of CNNs: Train Cohort 1 using Original and adversarial images and test on Original Cohort 2 images

| Approach | CNN 1: Ensemble of 7 CNNs | CNN 2 : Ensemble of 7 CNNs | CNN 3: Ensemble of 7 CNNs | Overall : Ensemble of 21 CNNs |
|---|---|---|---|---|
| FGSM | 78.05% (0.82 AUC) | 78.9% (0.82 AUC) | 79.32% (0.84 AUC) | 82.27% (0.88 AUC) |
| 1-pixel attack | 77.63% (0.8 AUC) | 78.05% (0.82 AUC) | 78.9% (0.82 AUC) | 81.43% (0.85 AUC) |

**Table 6**: Train Cohort 1 using Original and adversarial images and test on adversarial Cohort 2 images

| CNN 1 | | CNN 2 | | CNN 3 | |
|---|---|---|---|---|---|
| FGSM | 1-pixel attack | FGSM | 1-pixel attack | FGSM | 1-pixel attack |
| 68.35% | 73% | 65.77% | 73% | 70.04% | 73.4% |
| 68.35% | 71.3% | 67.5% | 70.88% | 66.24% | 72.15% |
| 67.5% | 72.5% | 66.66% | 71.72% | 65.8% | 73.83% |
| 67.08% | 71.3% | 66.66% | 73% | 69.62% | 71.72% |
| 67.5% | 70.04% | 67.93% | 72.15% | 69.19% | 73.41% |
| 67.93% | 72.5% | 67.5% | 73% | 69.19% | 73% |
| 67.93% | 73% | 67.93% | 72.5% | 68.35% | 73.41% |

**Table 7**: Ensemble of CNNs: Train Cohort 1 on original and adversarial images and test on Adversarial Cohort 2 images

| Approach | CNN 1: Ensemble of 7 CNNs | CNN 2 : Ensemble of 7 CNNs | CNN 3: Ensemble of 7 CNNs | Overall : Ensemble of 21 CNNs |
|---|---|---|---|---|
| FGSM | 73% (0.7 AUC) | 72.5% (0.69 AUC) | 74.68% (0.75 AUC) | 77.63% (0.81 AUC) |
| 1-pixel attack | 76.79% (0.79 AUC) | 78.05% (0.82 AUC) | 78.5% (0.82 AUC) | 80.16% (0.84 AUC) |

and one-pixel attack) for Cohort 1. Then the Cohort 1 adversarial images were divided into 70% training and 30% validation to perform data augmentation as mentioned earlier and the adversarial images were added to the originals of Cohort 1. CNN training was performed on the enhanced Cohort 1 dataset. Now the retrained CNNs were tested separately on original and adversarial Cohort 2 images. This analysis was conducted to check how much reduction (for Cohort 2 original images) or improvement of performance (for Cohort 2 adversarial images) was achieved. For FGSM, after training on an enhanced training set and test on original Cohort 2 images, we found that the performance of classification for all CNNs was reduced by 2-3%. Whereas we observed an improvement in classification accuracy of 15-20% after test on adversarial Cohort 2. For the one-pixel attack, after training on the enhanced training set and test on original Cohort 2, we found that the classification accuracy for all CNNs was reduced by 2-4%. Whereas we observed an improvement in classification accuracy of 1-3% after test on adversarial Cohort 2. Detailed results are shown in Table 4-7.

The results with adversarial images in training plus an ensemble of classifiers is more accurate than a single classifier on clean data. However, it is 13% less accurate than an ensemble on clean data for FGSM and 10% less accurate for a single pixel attack. Still, it is a big improvement over the 30% plus drop from FGSM seen without mounting a defense.

## 6 Conclusions

In summary, this paper addresses the problem of an adversarial attack on medical images with an application to lung nodule malignancy prediction. The paper also proposes an ensemble of CNNs as a defense strategy against the adversarial attacks. If a person intentionally or unintentionally changed a CT scan by adding some noise or changing a pixel value, then the deep learning algorithms may lose accuracy. In this paper, we studied two types of adversarial attack : FGSM and one-pixel attack. In terms of reducing the classification accuracy, we found that the FGSM attack was more successful than one-pixel attack. In this paper, we proposed a multi-initialization CNN ensemble approach. Each of CNN architecture was trained using seven different initializations. From each CNN, the pseudo probability was taken to create an ensemble by averaging. We found an improvement in malignancy prediction by our CNN ensemble approach. By adding adversarial images in the training set, we further minimized the effect of adversarial attack. As it is a preliminary study, 2-D slices were used, which was our study's limitation. Our study also used a semi-automatic approach to segmentation. In future we will work on a 3D adversarial attack and defense strategy.

## Acknowledgments

This research is partially supported by the NIH grants (U01 CA143062), (U24-CA180927), (U01-CA186145), (U01-CA196405), and (U01 CA200464), NSF under award number 1513126, by the State of Florida Dept. of Health under grant (4KB17).No conflicts of interest, financial or otherwise are declared by the authors.

## References

[1] A. C. Society, "Cancer facts & figures," *American Cancer Society*, 2016.

[2] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial examples in the physical world," *arXiv preprint arXiv:1607.02533*, 2016.

[3] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," *arXiv preprint arXiv:1312.6199*, 2013.

[4] R. Paul, S. Hawkins, M. B. Schabath, R. J. Gillies, L. O. Hall, and D. B. Goldgof, "Predicting malignant nodules by fusing deep features with classical radiomics features," *Journal of Medical Imaging*, vol. 5, no. 1, p. 011021, 2018.

[5] Y. Mirsky, T. Mahler, I. Shelef, and Y. Elovici, "Ct-gan: Malicious tampering of 3d medical imagery using deep learning," *arXiv preprint arXiv:1901.03597*, 2019.

[6] S. G. Finlayson, H. W. Chung, I. S. Kohane, and A. L. Beam, "Adversarial attacks against medical deep learning systems," *arXiv preprint arXiv:1804.05296*, 2018.

[7] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[8] M.-I. Nicolae, M. Sinn, M. N. Tran, B. Buesser, A. Rawat, M. Wistuba, V. Zantedeschi, N. Baracaldo, B. Chen, H. Ludwig, I. Molloy, and B. Edwards, "Adversarial robustness toolbox v0.10.0," *CoRR*, vol. 1807.01069, https://arxiv.org/pdf/1807.01069.

[9] J. Su, D. V. Vargas, and K. Sakurai, "One pixel attack for fooling deep neural networks," *IEEE Transactions on Evolutionary Computation*, 2019.

[10] D. Kondratyuk, "One pixel attack," 2018, https://github.com/Hyperparticle/one-pixel-attack-keras.

[11] S. Qiu, Q. Liu, S. Zhou, and C. Wu, "Review of artificial intelligence adversarial attack and defense technologies," *Applied Sciences*, vol. 9, no. 5, p. 909, 2019.

[12] X. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial examples: Attacks and defenses for deep learning," *IEEE transactions on neural networks and learning systems*, 2019.

[13] N. Akhtar and A. Mian, "Threat of adversarial attacks on deep learning in computer vision: A survey," *IEEE Access*, vol. 6, pp. 14 410–14 430, 2018.

[14] F. Tramèr, A. Kurakin, N. Papernot, I. Goodfellow, D. Boneh, and P. McDaniel, "Ensemble adversarial training: Attacks and defenses," *arXiv preprint arXiv:1705.07204*, 2017.

[15] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*. Springer, 2014, pp. 818–833.

[16] F. Chollet *et al.*, "Keras: The python deep learning library," *Astrophysics Source Code Library*, 2018.

[17] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[18] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural networks for machine learning*, vol. 4, no. 2, pp. 26–31, 2012.

[19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[20] A. Y. Ng, "Feature selection, l 1 vs. l 2 regularization, and rotational invariance," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 78.

[21] M. B. Schabath, P. P. Massion, Z. J. Thompson, S. A. Eschrich, Y. Balagurunathan, D. Goldof, D. R. Aberle, and R. J. Gillies, "Differences in patient outcomes of prevalence, interval, and screen-detected lung cancers in the ct arm of the national lung screening trial," *PloS one*, vol. 11, no. 8, p. e0159880, 2016.

[22] N. L. S. T. R. Team, "Reduced lung-cancer mortality with low-dose computed tomographic screening," *New England Journal of Medicine*, vol. 365, no. 5, pp. 395–409, 2011.

[23] S. Hawkins, H. Wang, Y. Liu, A. Garcia, O. Stringfield, H. Krewer, Q. Li, D. Cherezov, R. A. Gatenby, Y. Balagurunathan *et al.*, "Predicting malignant nodules from screening ct scans," *Journal of Thoracic Oncology*, vol. 11, no. 12, pp. 2120–2128, 2016.

[24] A. Definiens, "Developer xd 2.0. 4," *Reference Book*, 2012.

[25] K. Hajian-Tilaki, "Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation," *Caspian journal of internal medicine*, vol. 4, no. 2, p. 627, 2013.