Final Report:

Developing a deep learning based system to detect malicious tampering in medical imaging

By:

Niv Bar-on, I.D: 204351944, E-mail: nivba@post.bgu.ac.il

Supervisors:

Dr. Yaniv Zigel Yaniv Zigel

E-mail: yaniv@bgu.ac.il

Dr. Gal Ben-arie

E-mail: galbe@bgu.ac.il

Prop. Ilan Shelef

E-mail: IlanS@clalit.org.il

Department of Biomedical Engineering

Faculty of Engineering Science

Ben Gurion University of the Negev

Abstract

Recent progress in image and video manipulation are provoking a heated discussion about its dangers in a variety of areas such as media, politics and more. In this project we discussed and demonstrate the danger of abusing this progress for medical imaging manipulation. In addition, we present proof of concept for countermeasure against this danger.

As part of this project, we developed a novel system for injecting synthetic tumors into mammograms. The forgery capabilities of this system were tested with the help of four expert breast radiologists. We also developed an autoencoder-based forgery detection system and tested its performants on both seen and unseen forgery methods.

We hope that this project will advance the important discussion in dealing with deepfake in medical images and that the solution we have proposed will be a first step on the way to a more comprehensive countermeasure that will be widely assimilated in the medical systems.

Acronyms

Table 1:Acronyms Table

Acronym	Meaning
PACS	Picture Archiving and Communication Systems
NHS	National Health Service
GAN	Generative Adversarial Network
CGAN	Conditional Generative Adversarial Network
DW	Digital Watermarking
RONI	region of noninterest
ROI	region of interest
WGN	White Gaussian Noise
CNN	Convolutional Neural Network
BIRADS	Breast Imaging-Reporting and Data System
FT	ForensicTransfer
NPV	Negative Predictive Value
PPV	Positive Predictive Value
TNR	True Negative Rate
TPR	True Positive Rate
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
CC	Craniocaudal
MLO	Mediolateral Oblique

1. Introduction

Medical imaging is an indispensable component of modern medical practice [1]. Images retrieved from the different modalities are stored in the Picture Archiving and Communication Systems (PACS). The PACS acts as an efficient database that enables easy access to the images as well as to additional medical data [2]. PACS and similar digital systems has significantly streamlined the activity of healthcare organizations around the world but at the same time exposed them to many security breaches. Indeed, the healthcare system is one of the most targeted sectors for cyber-attacks, mainly because it's a soft target which is rich with valuable data [3]. On May 2017, as part of a global ransomware attack, over 600 organizations belonging to England National Health Service (NHS) were compromised. During the cyberattack, 46 hospitals were locked out of their digital systems and medical devices, resulting significant disruption across the NHS [4]. Though the attack did not directly target imaging centers and as far as we know such an attack did not happen yet, on 2019, Mirsky et al. demonstrate how one can gain access to medical imagery with a man-in-the-middle approach in order to inject and remove lesions causing misdiagnosis [5]. Such tasks like injecting and removing lesions from medical images became much more feasible thanks to advances in the field of machine and deep learning. One of the deep learning techniques which is utilized to synthetize realistic data is called Generative Adversarial Network (GAN) and it consists of two adversarial deep neural networks- Generator and Discriminator. The Generator network creates synthetic samples of given domain (mostly images), by passing a random noise through the net- mapping a random noise to an image. The Discriminator network receives the original samples and synthetic samples as an input and try to determine for each input whether it is original or synthesized. These two networks feed on each other and train simultaneously in a zero-sum game- the generative network trying to forge a sample that discriminative network will classify as real and the discriminative network try to detect as many fake samples as possible [6].

In this project we used Conditional GAN (CGAN), a different version of GAN since it learns the mapping from conditioned image to unconditioned image rather than from random noise to an image.

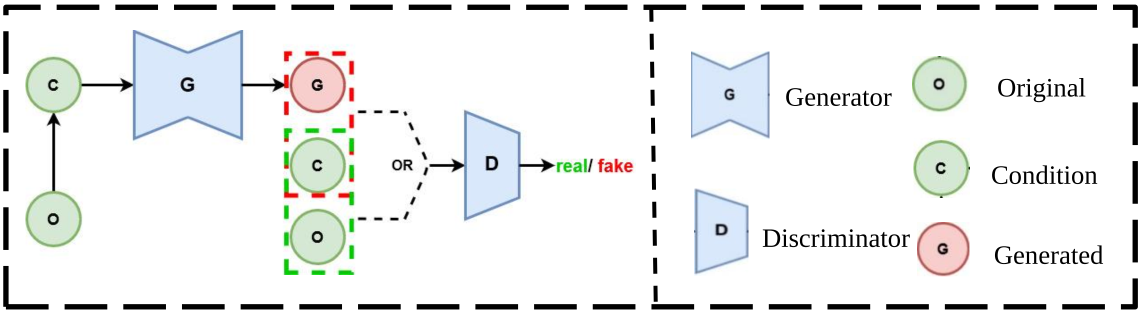

Figure 1: A schematic view of CGAN

By applying some condition to the original image (O) (for example zeroing a part of it), we create the condition image (C). Based on the conditioned image the Generator tries to create realistic unconditioned image (G), the Discriminator (D) gets the condition image with the original image or the generated image and tries to classify for real or fake [7].

With recent advances in machine learning and synthetic image rendering image manipulation has reached unprecedented levels of diffusion and sophistication and so, finding an algorithmic defense against image forgery remains an extremely open and challenging problem [8]. Contrary to many other images the medical images are usually uncompressed and maintain a very high resolution and imaging systems produce very different noise patterns than standard cameras [5]. It is, therefore, important to address forging of medical images individually.

Countermeasures against forgery of medical imagery were already proposed in the past, Among them is Digital Watermarking (DW) [9]. The main disadvantage of DW is that addition of data into a medical image decreases visual quality of the medical image and can cause false diagnosis [10]. To avoid masking vital information, it has been suggested to embed watermark on region of noninterest (RONI) of the medical image preserving region of interest (ROI) [11] , this solution will not prevent forgery because the forgery usually occur in the ROI. Few other additional methods were suggested to deal with forgeries in medical images [10] [12] Ant though demonstrating good results, these were only tested on a relatively simple forgery methods like copy-move forgery and splicing. Furthermore, none of the existing solutions have shown robustness to unseen forgery methods.

As far as we know, no countermeasure has yet been proposed against complex forgery of medical imagery (like GAN), which does not require data addition to the image and is robust to unseen forgery methods. In this project we propose an auto-encoder based countermeasure to address this highly important problem.

To create a repository of forge medical images to the current study we used healthy mammography images into which lesions were injected. We decided to use mammography since it is one of the most common imaging modalities being used [13]. Despite the brisk discussion in the literature regarding the effectiveness of mammography as a screening tool [14], it is accepted as such in most western countries [15].

1.1. Goals

This project has two main goals:

1. Develop a system for tumor injection in medical images- The system will allow us to create high-level counterfeiting of medical images, by synthesizing images of tumors and injecting them into mammograms. The system will be based on CGAN, one of the state-of-the-art synthesis methods today. The counterfeiting capabilities of this system will be tested using expert breast radiologists.
2. Develop a forgery detection system- This system should be able to deal with complex forgeries (like CGAN) without adding any data to the image and to show generalization capabilities against unseen forgeries methods. This System will be test on seen and unseen forgeries methods.

2. Methods

In this chapter we will explain in detail about the structure and components of both tumor injection system and forgery detection system. In addition, we will present the test methodology of the two systems and review the database we used.

2.1. Tumor injection

We used CGAN to create forgeries and inject tumors into healthy mammograms, in a process followed by Mirsky et al. [5]. The injection process combined two CGAN systems connected in a cascade, so that the first one produces a low-resolution image and the second produces a higher-resolution image from it, a similar idea has been proposed in the past and called progressive growing GAN [16] .

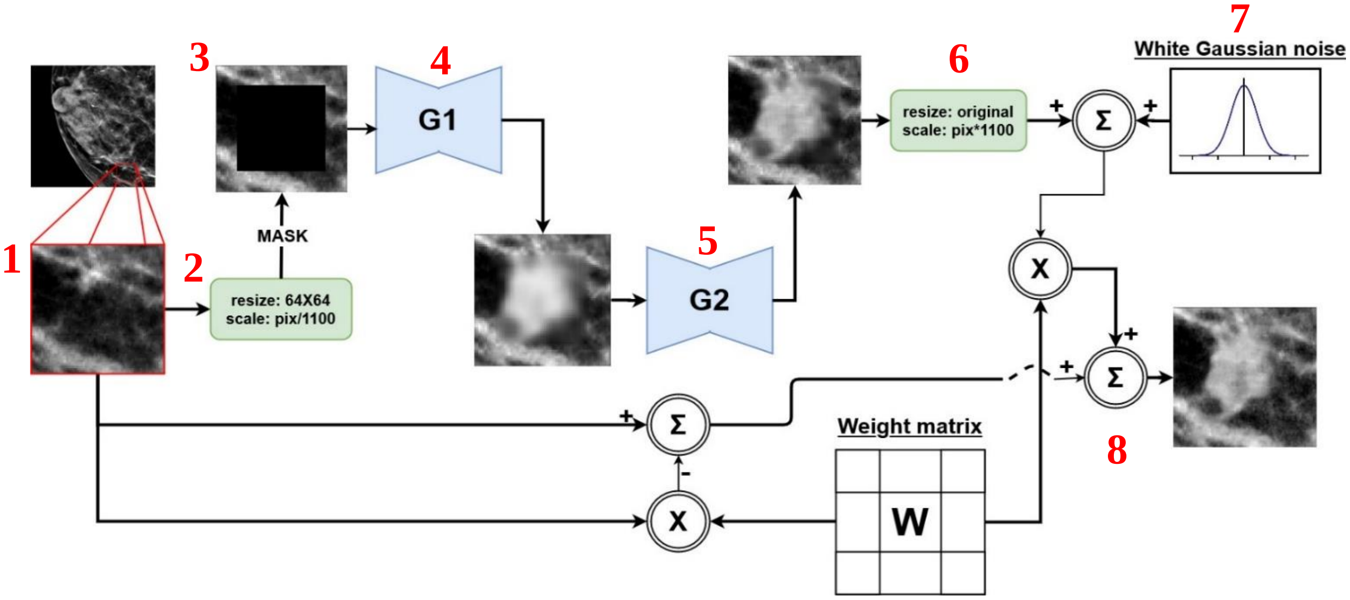

Figure 2: schematic view of Mammo-CGAN

Stage 1: Choosing a square within the mammogram in which we want to inject a tumor.

Stage 2: Resizing the square to be 64X64 pixels with bilinear interpolation and scale it by dividing the pixels value by 1100 (to reduce the pixels intensity dynamic range)

Stage 3: Conditioning the image by zeroing 44X44 pixels in the middle of it.

Stage 4: Feeding the condition Image to the first CGAN (CGAN1) Generator to create a blur tumor in the middle.

Stage 5: Feeding the first Generator output as a condition image to the second CGAN (CGAN2) Generator.

Stage 6: Resizing the square to its original size with bilinear interpolation and scale it back by multiplying the pixels value by 1100.

Stage 7: Adding White Gaussian Noise (WGN) to the image to hide the artifact of the resizing, the standard deviation of the noise chose by the user (default value- $\sigma = 10$). the result will be called **G**.

Stage 8: Merging to- **M** the generated image **G** with the original image **O** by weighted average using weight matrix **W** (eq 2.1):

$$\mathbf{M}(\mathrm{i,j}) = \mathbf{W}(\mathrm{i,j}) * \mathbf{G}(\mathrm{i,j}) + \left(1 - \mathbf{W}(\mathrm{i,j})\right) * \mathbf{O}(i,j), \quad \text{(eq 2.1)}$$

2.1.1. weight matrix

The values in the weight matrix range from 0 to 1, the matrix is calculated as function of the distance of the pixels from the center of the image (part A, in eq 2.2) and the intensity of the pixels in G (part B, in eq 2.2) in the following way:

$$\mathbf{W}(\mathrm{x,y}) = \underbrace{(1 - \mathbf{D}(\mathrm{x,y})^{p})}_{A} * \underbrace{\mathrm{S}(\frac{\mathbf{G}(\mathrm{x,y}) - Th}{\alpha})}_{B}, \quad \text{(eq 2.2)}$$

S- sigmoid function $\mathrm{S}(\mathrm{x}) = \frac{1}{1+\exp{(-x)}}$, (eq 2.3)

Th- threshold for the sigmoid function, chosen by the user (default value- 400)

α- smoothing parameter chosen by the user (default value- 20)

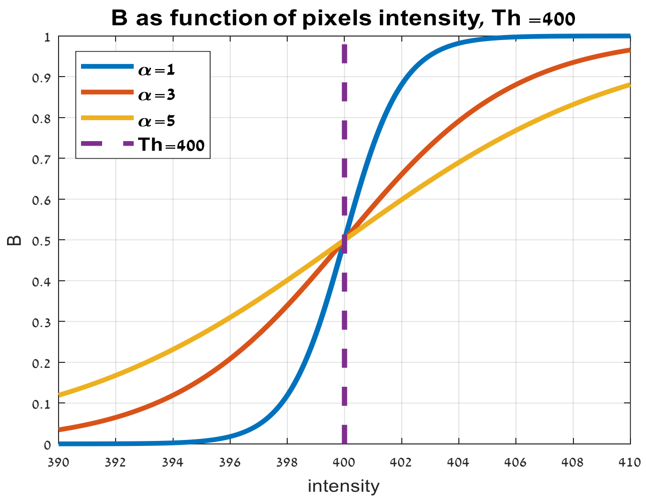

Figure 3: B part as function of intensity for variable alphas

D: distance matrix grows linearly between 0 to 1 for instance:

$$\begin{pmatrix} 2 & 2 & 2 & 2 & 2 \\ 2 & 1 & 1 & 1 & 2 \\ 2 & 1 & 0 & 1 & 2 \\ 2 & 1 & 1 & 1 & 2 \\ 2 & 2 & 2 & 2 & 2 \end{pmatrix} \cdot \frac{1}{2}$$

P: parameter that controls A parameter that controls the descending concavity degree of the distance depending part, chose by the user (default value- 3)

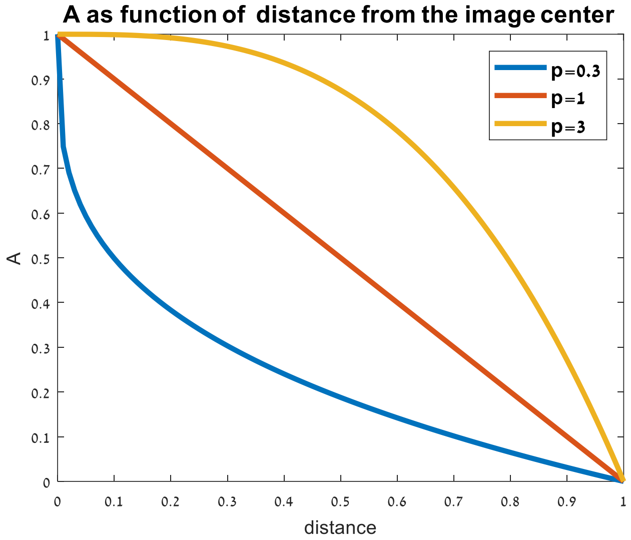

Figure 4: part A as function of distance with variable p

2.1.2. CGANs architecture and training

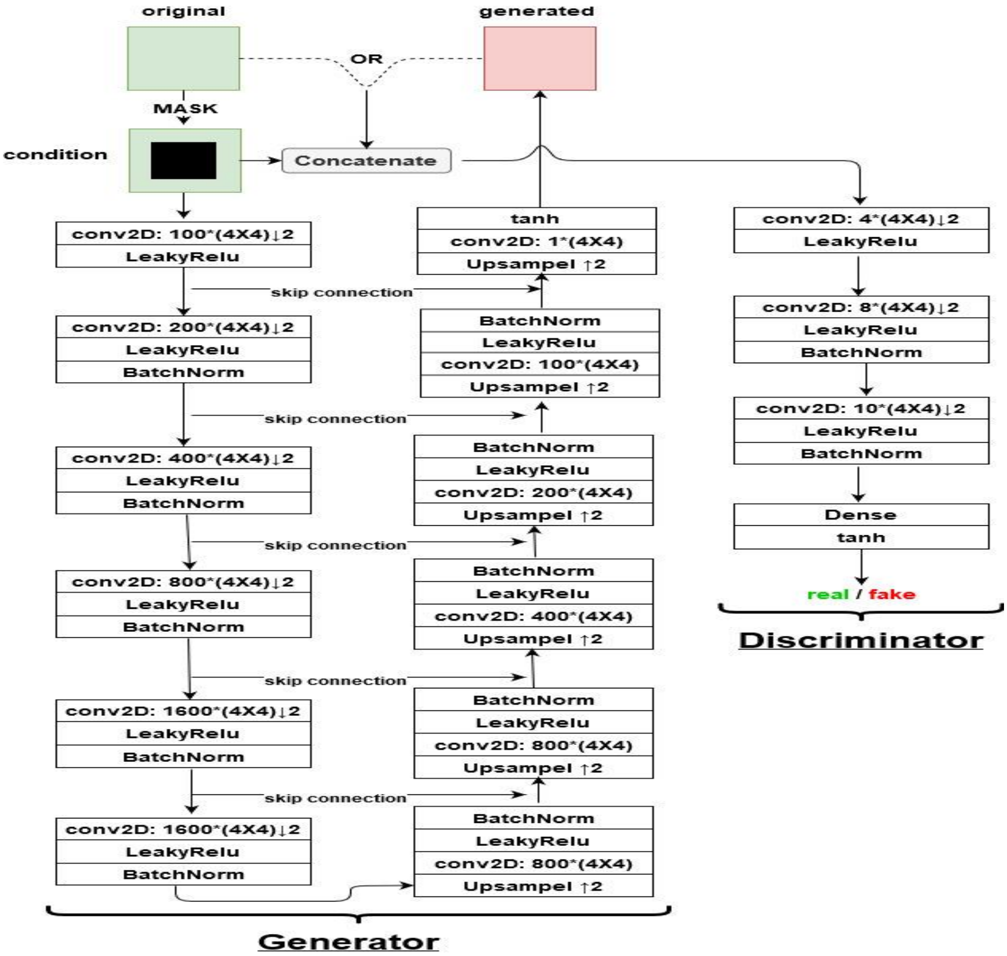

Figure 5: CGAN1 architecture

Two CGAN architectures use 2 different neural network- Generator and Discriminator. The Generator architecture, in our implementation, as a U-net like architecture [17] consists of a contracting path (left side) and expansive path (right side). The layers in the contracting path uses 2-dimensional convolution with stride of two (↓2) for down sampling. Every step in the expansive path consists of an up sampling of the feature (↑2) map, 2-dimensional convolution with stride of one, and a concatenation with the correspondingly cropped feature map from the contracting path. Each layer on both paths uses LeakyRelu as activation function except of the last one in the expansive path that use tanh. In each layer except of the first and the last one we used batch normalization.

The Discriminator architecture is a simple Convolutional Neural Network (CNN) all the layer except of the last one preforms 2-diamentional convolution followed by

LeakyRelu activation function. The last layer is a Dense layer with one neuron followed by tanh activation function. All layers except the first and the last one uses batch normalization. The purpose of the Discriminator is expressed only in the training phase and is to serve as an adversarial network to the Generator, so after the training phase we will no longer use it

The exact architecture of CGAN1 is described in figure 5. CGAN2 have a similar architecture to CGAN1 with minor changes in the layers composition (averrable in the source code).

Both models trained on 2880 images augmented from 36 images of mammography tumors annotated by a specialist radiologist and optimized with Adam optimizer [18]. We will emphasize the fact that both models were trained on only unhealthy images and therefore learned to synthesize only unhealthy images.

2.1.3. Testing the injection system

To test for the authenticity of the fake images we used four experience breast radiologists (at least 3 years of experience of breast radiology). We first presented 80 mammograms which included 45 fake images, 20 real healthy mammogram (BIRADS 1) and 15 real unhealthy mammograms (BIRADS ≥3). The radiologists were not told that some of the images were tampered and they were asked to classify the mammograms as healthy or unhealthy and to mark lesions within the mammograms. For each case and for the whole series a space was given to add any remark regarding the images.

In the following part the radiologists were told that some of the mammogram may have been tampered and were asked to classify the mammograms as healthy or unhealthy, real or fake and to mark lesions within the mammograms. For this stage, the images contained 21 mammograms in the following mix: 7 fakes (created by injecting a tumor into a healthy mammogram), 7 reals healthy mammogram (BIRADS 1) and 7 real unhealthy mammograms (BIRADS ≥3).

2.2. Forgery detection

In order to develop a robust countermeasure against advanced forgeries (Like GAN), we implemented a system called ForensicTransfer (FT) [8], a weekly supervised CNN that shows great generalization and adaptation capabilities. Despites FT good performance we could not implement it as is for the following reasons:

- Medical images maintained much higher resolution then regular images, FT was tested on lower resolution images.
- The relative proportion of forgery in a medical image can be small and still cause misdiagnosis.
- Medical images from same modality can have different image sizes and CNNs are usually trained to deal with fixed resolution.

To overcome these gaps, we developed the following system:

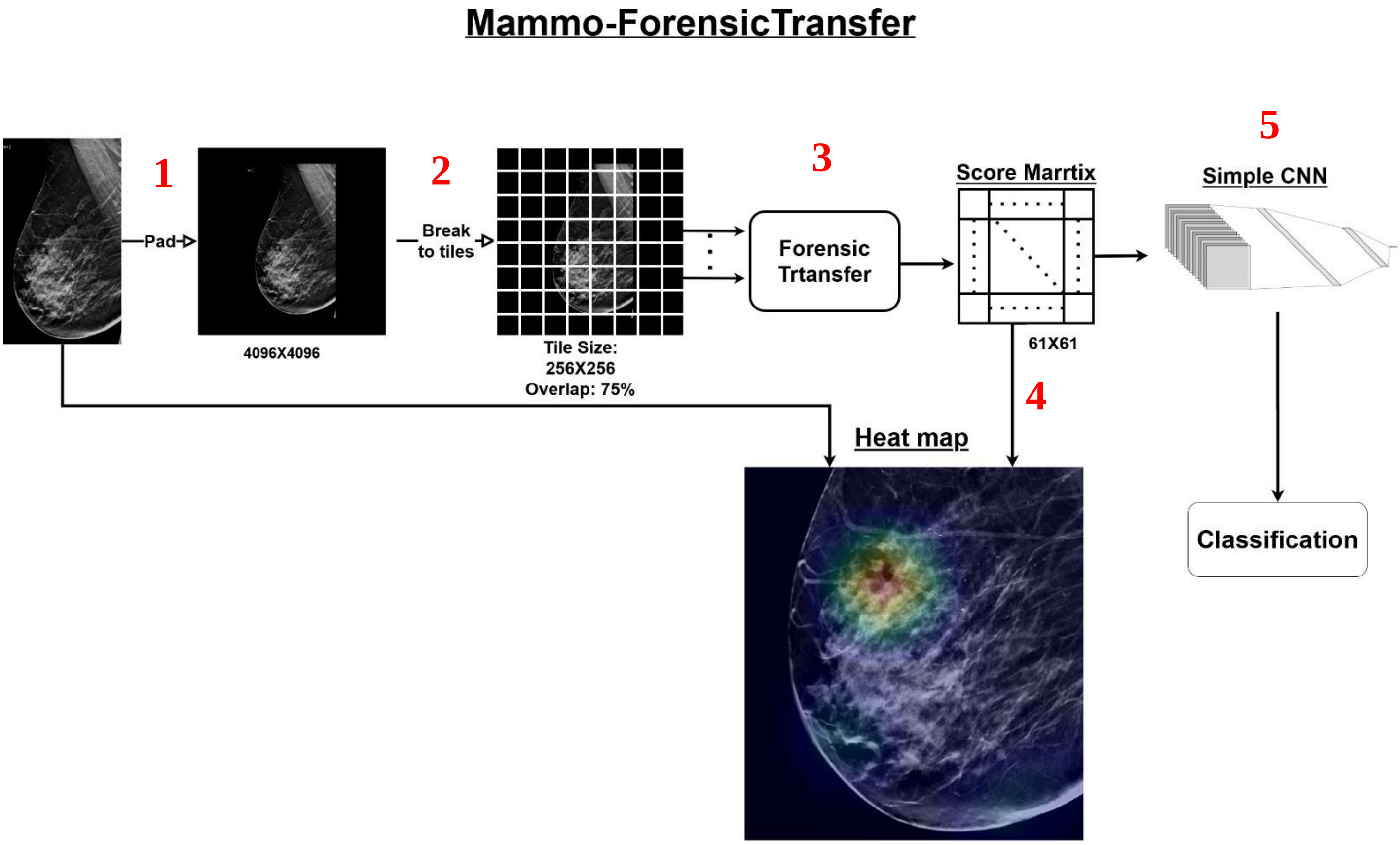

Figure 6: Mammo-ForensicTransfer schematic view

<u>Step 1:</u> Padding the image with zeros to be 4096X4096 pixels- to deal with non-fixed image size.

<u>Step 2:</u> Breaking the image into 256X256 size tiles with overlap of 75%

Step 3: Sending any tile that is not fully black (determined using simple threshold) to FT. The network returns a score ranging from 0 to 1 for how fake the tile looks. All scores are maintained in a 61X61 scores matrix.

Step 4: Using the scores matrix to create a heat map, by allaying a 4X4 average matrix and bilinear interpolation to the image. "Hot" areas in the heat map represents areas where there is likely to be a forgery in.

Step 5: Feeding the score matrix into a simple CNN- consists of one convolutional layer 1 pooling layer and three fully connected layer (averrable in the source code). This CNN returns a score between 0 to 1 for how fake the Image looks, if the score is higher than 0.5, we will classify the image as fake.

2.2.1. ForensicTransfer

FT is an auto-encoder-based [19] forgery detection system which aims to maintain a good generalization to unseen forgery methods. In this system, the real/fake decision is made based on information drawn from the latent space that constrained to preserve all the data necessary to reconstruct the image in compact form. Therefore, the latent space holds both the image representation and the data used for the real/fake decision, but these pieces of information live in orthogonal spaces, and do not interfere with one another. This is obtained by dividing the latent space in two parts, one activated exclusively by real samples, and the other by fake samples. Since the network has to reproduce the image anyway, all relevant information on the input image is stored in both parts [8].

The auto-encoder learns an encoding function E(.) to map image $\mathbf{X}$ to latent vector representation $\mathbf{h}$ and a decoding function D(.) to provide an approximate reconstruction of the image- $\widehat{\mathbf{X}}$. The latent vector has 128 elements and conceptually divided into two parts: $\mathbf{h_0}$- consists of the first 64 elements of h and associate with real images and $\mathbf{h_1}$ - consists of the second 64 elements of h and associate with fake images.

During training, the degree of activation of the latent vector parts is checked by the l1 norm and a score is calculated regarding how fake the image look like in the following way: $score = \frac{\|\boldsymbol{h_1}\|_1}{\|\boldsymbol{h_0}\|_1+\|\boldsymbol{h_1}\|_1}$, Depending on the score the image is classified as real or fake: $c = \begin{cases} 1; & score > 0.5 \\ 0; & else \end{cases}$. Once the classification has been made a new latent vector $\mathbf{h}'$ is determined so that $\mathbf{h}'_{\mathbf{c}} = \mathbf{h}_{\mathbf{c}}$ and $\mathbf{h}'_{\mathbf{1-c}} = 0$. Then using the decoder, a reconstruction been made $\widehat{\mathbf{X}} = D(\mathbf{h}')$.

The loss function is a weighted sum between two losses: $L = 5 \cdot L_{ACT} + L_{REC}$. L_{ACT} is an activation loss and determined by l2 norm between the true label and the score: $L_{ACT} = (true\ label - score)^2$. L_{REC} is a reconstruction loss and determined as l1 norm between $\mathbf{X}$ and $\widehat{\mathbf{X}}$: $L_{REC} = \frac{1}{256^2}\sum_{i=0}^{255}\sum_{j=0}^{255}\left|\mathbf{X}(i,j) - \widehat{\mathbf{X}}(i,j)\right|$. Note that L_{REC} is taking into account for both real and fake images and so the network must maintain sufficient amount of data to reconstruct the image anyway, preventing the network from learning only the relevant information in order to identify the forgery method the network trained on and thus improves its generalization capability.

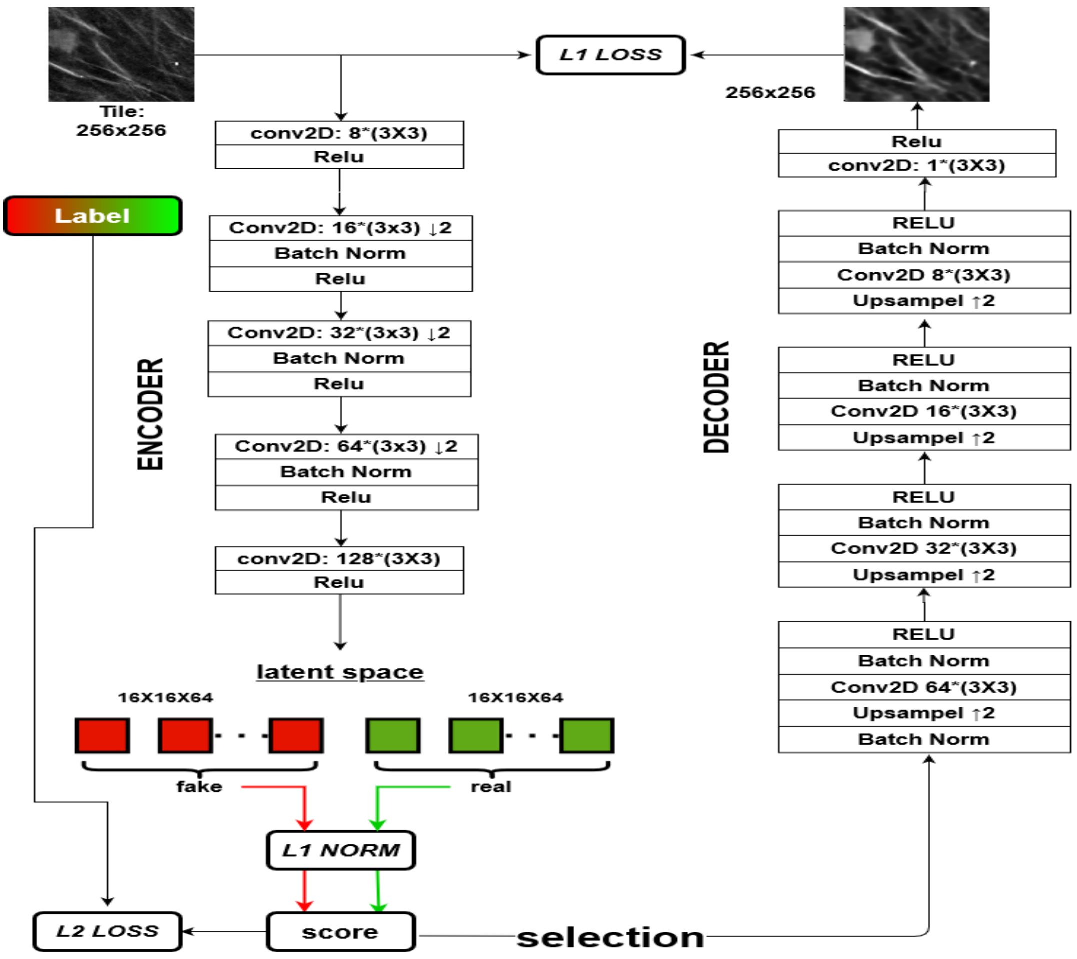

Figure 7:ForansicTransfer architecture

2.2.2. Testing the forgery detection system

We examined the detection system performance on two different data sets - the first one called Source domain, consists of 101 images (52 fakes and 49 reals) that were also given to the radiologists at the stage of examination of the tumor injection system. The forgeries created by the same method on which the system has trained on- created by injecting lesions into healthy mammograms using our tumor injection system.

The other set consists of forgeries created by unseen forgery method and called target domain. The target domain is used to test the system generalization capabilities.

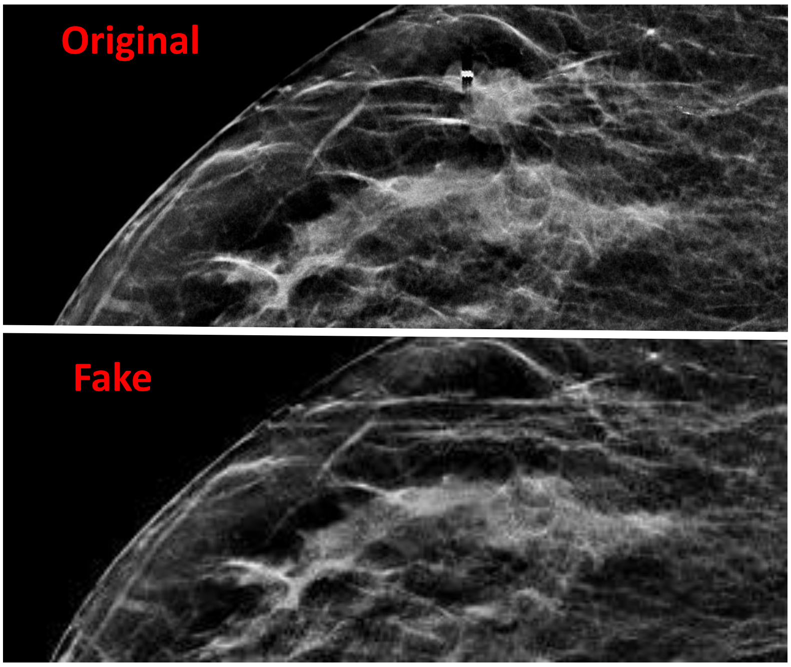

Figure 8: Target domain forgery example

The Target domain consists of 64 images- 32 fakes that created by removing existed lesions from mammograms using Nvidia inpainting tool [20] and 32 images that just went through NVIDIA's system but have remained untouched.

2.3. Database

All the data for this project are mammography images with resolution of 2457X1996 pixels, taken from Soroka Hospital's image repository. The database is divided into 5 separate data sets:

Tumor injection system training set- consist of 2880 images augmented from 36 images of mammography tumors annotated by a specialist radiologist. Used to train CGAN1 and CGAN2.

Forgery detection training set- consists of 594 mammograms in the following mixture: 494 forgeries created by injecting tumor into healthy mammograms, using our tumor injection system and 100 real untampered mammograms. Used to train the forgery detection system.

Forgery detection validation set- consists of 28 mammograms in the following mixture: 18 forgeries created by injecting tumor into healthy mammograms, using our tumor injection system and 10 real untampered mammograms. Used to optimize the training process of the forgery detection system by avoid overfitting and hypermeter tuning.

Source domain testing set- consists of 101 mammogram images in the following mixture: 52 forgeries created by injecting tumor into healthy mammograms, using our tumor injection system and 49 real untampered mammograms. Used to assess the tumor injection forgery system using four experience breast radiologists and to test the forgery detection system.

Target domain testing set- consists of 64 mammogram images in the following mixture: 32 forgeries created by removing lesions from unhealthy mammograms, using NVIDIA inpainting tool and 32 real untampered mammograms. Used to test the forgery detection system generalization capability.

3. Results

This chapter presents the products of the systems we have developed and their test results.

3.1. Tumor injection

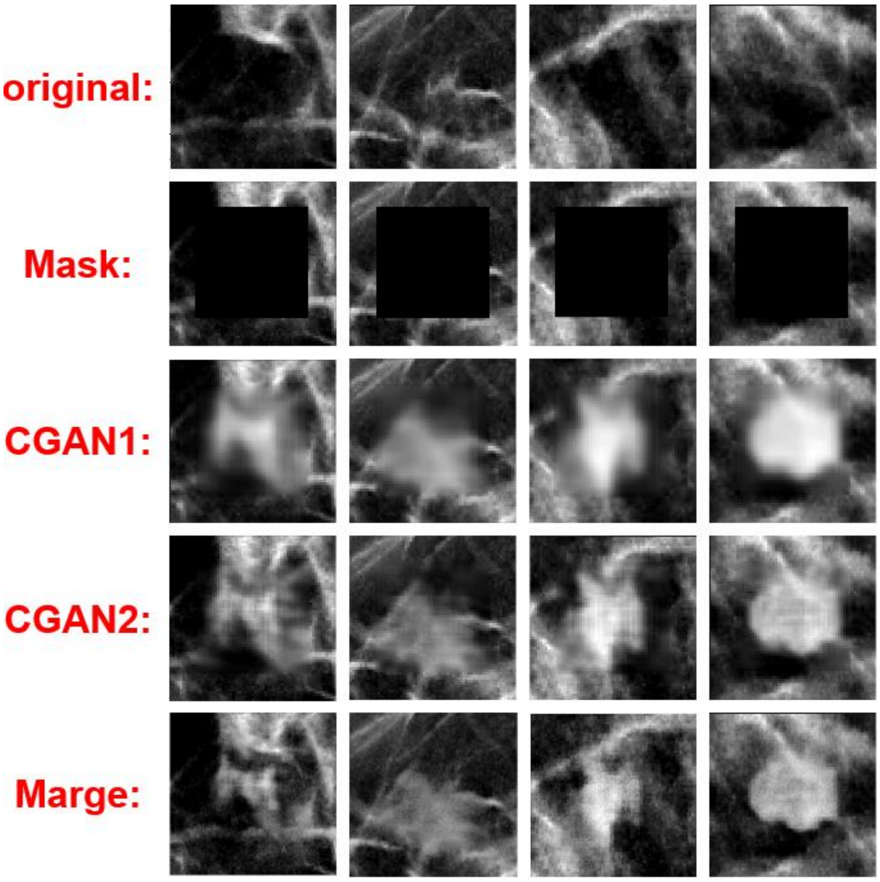

Figure 9: Pictures of the process of injecting the tumors into healthy tissue

In this figure you can see the stages of tumor injection. Every column is a different example of the process. The first row consists of a healthy tissue images taken from mammograms. The second row is the same images masked in the middle- input of CGAN1. The third row is the output of CGAN1- input of CGAN2. The fourth row is the output of CGAN2. The fifth row is the final result- weighted average of the output of CGAN2 Contained with white gaussian noise with the original image. An extension of the injection stages you can find in the methods section.

As mentioned, the fake images authenticity was tested using 4 expert breast radiologists. In the blinded part we intended to check how much the injection operation might cause a misdiagnosis of a healthy image as an unhealthy image. To this end, we summarized the labels given by the four radiologists to the original healthy images and the fake images.

Table 2: Labeling of real-healthy and fake images as given by the radiolegists

labels classes	**unhealthy**	**healthy**	Total
Real healthy	36	44	80
Fake	124	56	180
Total	160	100	260

In Table 2 you can see the radiologist's classifications of real-healthy images and fake images (created by injecting tumor into a healthy images) to healthy and unhealthy. In order to test whether injecting a tumor into a healthy image increases the chance to an error of type - classifying a healthy image as unhealthy image, we performed a Chi squared test on the table data. Our null hypothesis was that the prevalence of this type of error is the same for real-healthy images and fake images. Using Chi test we succeed to contradict the null hypothesis with significance of 0.026%. To assess the power of the statistical test we used permutations- under the assumption that the error on the real healthy images is distributed $Bin(\frac{36}{80})$ and the error on the fake images is distributed $Bin(\frac{124}{180})$, we repeatedly sampled the table data and preformed a Chi squared test on each sample. In this process we have reached a statistical significance (p<5%) in 96.8% of the experiments.

In the open part of the experiment we wanted to check how convincing our fakes are. The Radiologists were told that some of the images may be fake and they were asked to classify every image as real or fake.

Table 3: confousion matrix of the radiolegists fake image detaction cepabilitys

		Actual class		
		real	**fake**	
Predicted class	**real**	**48**	**15**	NPV: **76.2%**
	fake	**8**	**13**	PPV: **61.9%**
		TNR: **85.7%**	TPR: **46.4%**	Acc: **72.6%**

In Table 3 we presrent a summery of the radiolegists fake image ditaction cepabilitys, in this matrix the positive class is fake and the negative class is real. We also calculated a few other metrics for the classification ability of the radiologists:

- NPV- Negative Predictive Value calculated as follows: $NPV = \frac{TN}{TN+FN}$.
- PPV- Positive Predictive Value, also known as precision, calculated as follows: $PPV = \frac{TP}{TP+FP}$.
- TNR- True Negative Rate, also Known as specificity, calculated as follows: $TNR = \frac{TN}{TN+FP}$.
- TPR- True Positive Rate, also Known as sensitivity/recall, calculated as follows: $TPR = \frac{TP}{TP+FN}$.
- Acc: Accuracy, calculated as follows $Acc = \frac{TP+TN}{TP+FP+TN+FN}$.

(TP- true positive, TN- true negative, FP- false positive, FN- false negative).

3.2. Forgery detection

The forgery detection system performance evaluated on two data sets - Source domain and Target domain.

3.2.1. Source domain

This data set consists of the images given to the radiologists on both parts of the experiment (blind part and open part).

Table 4: confusion matrix of the forgery detection system performens on the Source domain.

		Actual class		
		real	**fake**	
Predicted class	**real**	**46**	**7**	NPV: **86.8%**
	fake	**3**	**45**	PPV: **93.8%**
		TNR: **93.9%**	TPR: **86%**	Acc: **90.1%**

Table 4 summarize the performance of the forgery detection system on mammograms from the Source domain.

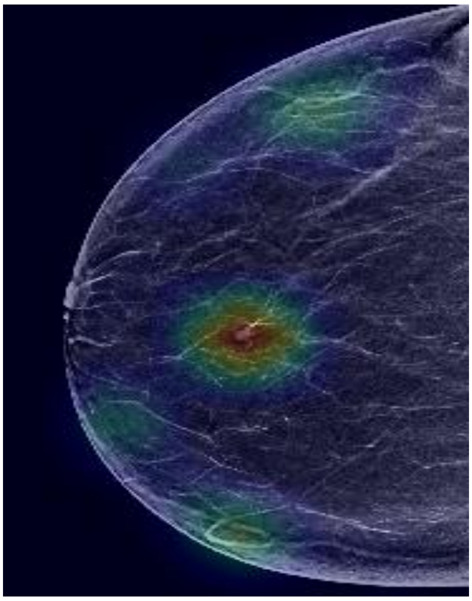

Figure 10: Heat map example.

Aside from classifying the image as real or fake, the forgery detection system produces a heat map. The red area in the heat map represents an area where the system estimates that the image is likely to be fake.

3.2.2. Target domain

The forgery detection generalization capabilities are tested on unseen forgery method called Target domain.

Table 4: confusion matrix of the forgery detection system performens on the Target domain.

		Actual class		
		real	**fake**	
Predicted class	**real**	**24**	**16**	NPV: **60%**
	fake	**8**	**16**	PPV: **66.7%**
		TNR: **75%**	TPR: **50%**	Acc: **64.5%**

Table 4 summarized the performance of the forgery detection system on mammograms from the Target domain.

4. Discussion

In this experiment we managed to show that injecting a tumor into a mammogram using our injection system increases the chance of classifying a healthy mammogram as unhealthy. We thus conclude that our forgeries do indeed look like real tumors. Even when radiologists were told that some of the images could be fake, they were able to detect only 46.4% of the fakes with total accuracy of 72.6%. This fact illustrates the deceptive potential of the system we have created for tumor injection and the need for a forgery detecting system.

Our forgery detection system shows great results detecting forgeries creating with Source domain forgery method, with total accuracy of 90.1%, showing better results than the expert radiologists on the same data set. Our forgery detection system also shows some generalization capabilities on Target domain with total accuracy of 64.5%. Despite this there is still a lot of room for improvement in this subject. We believe that training the detection system based on a more diverse database (in terms of forgery methods) will significantly improve system performance on unseen forgery methods. These results demonstrate that it is possible to create countermeasure against advanced forgery methods and even unseen forgery method in medical images.

4.1. Limitations

The experiment that included the radiologists suffered from a methodological problem- usually breast radiologists gets two views of the breast: Craniocaudal (CC) and Mediolateral Oblique (MLO). Injecting a tumor into two views of the same breath convincingly is a much more difficult task than injecting a tumor into a single view. Therefore, in our experiment the radiologists received only one view, which explains the larger-than-usual error rate in analyzing real mammograms.

As mentioned before, the Target domain data set has created using Nvidia's inpainting tool. Although this tool did an excellent job removing tumors from mammograms, one major issue arose in its use: when upload an image to NVIDIA's system the image is automatically resize and cropped to 512X512 pixels causing a significant information loss. we believe this issue has affected system performance on Target domain.

5. Conclusions

In this project we investigated a defensive solution against malicious tempering in medical images. For this end we developed, trained, and tested two novel systems. The first system is a GAN based-tumor injecting system for mammographic scans, which allowed injection of real-looking synthetic tumors into mammogram images. The second one is a forgery detection system- an autoencoder based system that aims to detect forgeries in mammogram images. The system presented showed great performance detecting forgeries from Source domain and some basic generalization capabilities on Target domain although there is room for improvement in this area.

We believe that forgery detection is an arms race, as time goes on both the forgers and those who try to prevent them getting better. For that reason, a solution to the problem of counterfeiting cannot be permanent and should get better and more sophisticated. The solution should be dynamic and vary according to technological progress. In order to maintain such a solution, it is important to promote active research on the subject.

Moreover, the systems we developed has other interesting applications as tumor injection system can be used as an augmentation tool, with the help of this system we can create and annotated data quickly and conveniently for training other learning systems. In addition, it will be interesting to test whether the system that detects forgeries can detect real tumors given a suitable data set for training.

6. References

[1] E. Samei, *Hendee's physics of medical imaging*, Fifth edition. Hoboken, NJ Chicester, UK: Wiley Blackwell, 2019.

[2] H. K. Huang, "Medical imaging, PACS, and imaging informatics: retrospective," *Radiol. Phys. Technol.*, vol. 7, no. 1, pp. 5–24, Jan. 2014.

[3] G. Martin, P. Martin, C. Hankin, A. Darzi, and J. Kinross, "Cybersecurity and healthcare: how safe are we?" *BMJ*, vol. 358, Jul. 2017.

[4] S. Ghafur, S. Kristensen, K. Honeyford, G. Martin, A. Darzi, and P. Aylin, "A retrospective impact analysis of the WannaCry cyberattack on the NHS," *Npj Digit. Med.*, vol. 2, no. 1, pp. 1–7, Oct. 2019.

[5] Y. Mirsky, T. Mahler, I. Shelef, and Y. Elovici, "CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning," *ArXiv190103597 Cs*, Jun. 2019, [Online]. Available: http://arxiv.org/abs/1901.03597. [Accessed: Nov. 14, 2019].

[6] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, vol. 3, pp. 2672–2680, 2014.

[7] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 5967–5976, Jul. 2017.

[8] D. Cozzolino, J. Thies, A. Rössler, C. Riess, M. Nießner, and L. Verdoliva, "ForensicTransfer: Weakly-supervised Domain Adaptation for Forgery Detection," *arXiv.org*, 2018. [Online]. Available: http://search.proquest.com/docview/2151583463/?pq-origsite=primo. [Accessed: Nov. 14, 2019]

[9] A. K. Singh, B. Kumar, G. Singh, and A. Mohan, *Medical Image Watermarking Techniques and Applications*, 1st ed. 2017. Cham: Springer International Publishing, 2017.

[10] G. Ulutas, A. Ustubioglu, B. Ustubioglu, V. Nabiyev, and M. Ulutas, "Medical Image Tamper Detection Based on Passive Image Authentication," *J. Digit. Imaging*, vol. 30, no. 6, pp. 695–709, 2017.

[11] S.C. Rathi and V.S. Inamdar, “ANALYSIS OF WATERMARKING TECHNIQUES FOR MEDICAL IMAGES PRESERVING ROI,” *Comput. Sci. Inf. Technol*, 2012.

[12] A. Ghoneim, G. Muhammad, S. U. Amin, and B. Gupta, “Medical Image Forgery Detection for Smart Healthcare,” *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 33–37, 2018.

[13] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2018,” *CA. Cancer J. Clin.*, vol. 68, no. 1, pp. 7–30, 2018.

[14] H. G. Welch, P. C. Prorok, A. J. O’Malley, and B. S. Kramer, “Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness,” *N. Engl. J. Med.*, vol. 375, no. 15, pp. 1438–1447, Oct. 2016.

[15] Melissa Conrad Stöppler, “Mammograms: Breast Cancer Diagnostic Screening Guidelines,” *MedicineNet*. https://www.medicinenet.com/mammogram/article.htm (accessed Nov. 14, 2019).

[16] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive Growing of GANs for Improved Quality, Stability, and Variation,” 20171027, [Online]. Available: http://arxiv.org/abs/1710.10196. [Accessed: Aug. 15, 2020].

[17] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” 20150518. [Online]. Available: http://arxiv.org/abs/1505.04597. [Accessed: Aug. 17, 2020].

[18] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *ArXiv14126980 Cs*, Jan. 2017. [Online]. Available: http://arxiv.org/abs/1412.6980. [Accessed: Sep. 09, 2020].

[19] M. Tschannen, O. Bachem, and M. Lucic, “Recent Advances in Autoencoder-Based Representation Learning,” *arXiv.org*, 2018. [Online]. Available: http://search.proquest.com/docview/2155543862/?pq-origsite=primo. [Accessed: Aug. 19, 2020].

[20] G. Liu, F. Reda, K. Shih, W. Ting-Chun, A. Tao, and B. Catanzaro, “Image Inpainting for Irregular Holes Using Partial Convolutions,” *arXiv.org*, 2018, [Online]. Available: http://search.proquest.com/docview/2072055200/?pq-origsite=primo. [Accessed: Sep. 01, 2020].