Universitatea
Transilvania
din Brașov

ȘCOALA DOCTORALĂ INTERDISCIPLINARĂ

Facultatea: Inginerie Electrică și Știința Calculatoarelor

Ing. Anamaria VIZITIU

Învățarea automată pentru generarea de imagini medicale, diagnosticare neinvazivă și asigurarea confidențialității datelor

Deep Learning for medical image generation, non-invasive diagnosis and privacy preservation

REZUMAT / ABSTRACT

Conducător științific

Prof.dr.ing. Florin MOLDOVEANU

BRAȘOV, 2020

Universitatea
Transilvania
din Brașov

D-lui (D-nei) .....

Componența

Comisiei de doctorat

Numită prin ordinul Rectorului Universității Transilvania din Brașov

Nr. ..... din .....

PREȘEDINTE:	- Prof. dr. ing. MORARU Sorin-Aurel Director de departament Universitatea Transilvania din Brașov
CONDUCĂTOR ȘTIINȚIFIC:	- Prof. dr. ing. MOLDOVEANU Florin Dumitru Universitatea Transilvania din Brașov
REFERENȚI:	- Prof. dr. ing. MICLEA Liviu Universitatea Tehnică din Cluj Napoca
	- Prof. dr. ing. POPESCU Dan Universitatea Politehnica din București
	- Prof. dr. ing. SUCIU Constantin Universitatea Transilvania din Brașov

Data, ora şi locul susținerii publice a tezei de doctorat: ....., ora ....., sala .....

Eventualele aprecieri sau observații asupra conținutului lucrării vor fi transmise electronic, în timp util, pe adresa anamaria.vizitiu@unitbv.ro

Totodată, vă invităm să luați parte la şedința publică de susținere a tezei de doctorat.

Vă mulțumim.

Contents

1. Introduction

1.1 Towards Data-Driven Medicine: Overview, Challenges and Future

Medical imaging is a gold-standard non-invasive procedure widely used in the diagnostic process. It produces detailed images of the internal structures of the human body for clinical purposes, such as disease prevention, diagnosis, treatment planning, and monitoring.

Nowadays, medical imaging has become an increasingly important part of the healthcare domain by facilitating early detection and improving patient well-being. The accuracy of the diagnosis depends to a large extent on the quality of the acquired image but, more importantly, on the interpretation of medical images. In terms of healthcare physicians' interpretation, the process is highly complex and prone to error due to the physicians' experience, subjectivity, and the level of fatigue. To these is added the extensive variations that exist between patients, the imaging complexity, and the heavy workload.

Over the past couple of years, due to technological progress and driven by the necessity to improve efficiency in the healthcare system, much effort has been invested in providing solutions for automated medical image interpretation. Although in practice, the medical imaging interpretation relies heavily on the physicians' expertise, with the experience and years of practice, this process becomes automatic to a certain extent. Thus, medical image analysis can very well be formulated for computer programs. Combining the power of computer science with the clinicians' knowledge to tackle the interpretation problem in medicine is a crucial step in improving the decision-making process.

Consequently, Artificial Intelligence (AI) for medicine has been introduced as a fast-growing computer science field with the potential to increase the likelihood of using automatic systems in clinical practice in the near future. Although AI implies cognitive computing, current technologies are far from achieving the ideal intended intelligence. In the best case, AI provides algorithmic means to understand, detect, and recognize patterns in large datasets. Machine Learning (ML) is the application of AI that relies upon data mining and statistics to enable computers to learn from vast amounts and various forms of data without being explicitly programmed [1].

The ability to extract meaningful insights from large datasets make ML-based solutions well-suited to solve complex data analysis problems, including medical image-based analysis. Implementing and adopting ML in healthcare-related problems presents multiple challenges. Since patient data have to be collected and manually annotated by medical experts, gathering enough suitable data and annotations remains one of the most biggest barriers to ML integration in the healthcare industry.

Despite being one of the fastest-growing sectors in the global economy, with the market being expected to reach $6.6 billion by 2021 [2], the biomedical industry is one of the main laggards in the adoption of data-driven solutions and practices. The gap takes place in an area that produces extremely abundant collections of valuable data. While biomedical data are abundant, they are challenging to circulate due to ethical and legal constraints.

Medical AI-systems have demonstrated their ability to improve clinical decisions, increase patient safety, and reduce costs, but they are still virtually absent from day-to-day clinical care, as data to develop and train them exist, but are locked inside hospitals firewalls. With the currently adopted regulations, such as the European Union's General Data Protection Regulation (GDPR), several concerns have been raised about data privacy, security, and sharing. Such regulations restrict the use

or disclosure of personal data, which clearly affects the usage and development of data-driven algorithms. Consequently, as access to sensitive data is required in medical ML-based applications, there is an urge to address the data privacy and security concerns in a way that enables the share of patient-related health information while ensuring the progress and utility of ML in the healthcare sector.

In recent years, there has been a continuous effort invested in providing solutions that rely on AI for fast, accurate, and secure processing of medical data for screening, prevention, diagnosis, and personalized treatment guidance. However, certain challenges remain to be addressed to fully embrace these solutions in clinical practice.

1.2 Aims of the Thesis

With the rapid advances in AI technologies, the increased computing power, e.g., graphical processing units (GPUs) and cloud computing systems, and the abundance of data, the potential of AI techniques have begun to be studied in the context of medical data. The traditional clinical workflow of medical imaging interpretation consists of three separate stages: (i) raw data acquisition, (ii) reconstruction of the internal structure of the human body upon the acquired data, and (iii) clinical interpretation of the reconstructed image. However, in combination with ML-based solutions for clinical decision-making, a new stage, responsible for patient data privacy, integrity, and security, is required, as shown in Figure 1.1.

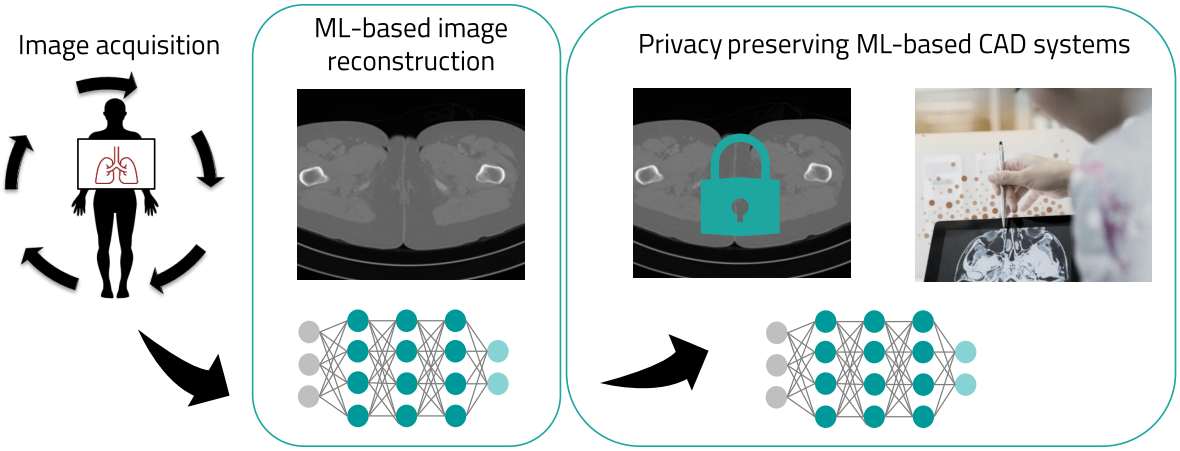


Figure 1.1: The process of artificial intelligence-based radiological image interpretation, from data acquisition to image-based diagnosis.

Given that radiology plays a major role in disease diagnosis and treatment, and knowing that imaging analysis is time-consuming and susceptible to human-level errors, the focus lies on investigating the potential benefits of machine learning in the field of medical data, with emphasis on radiology imaging. Hence, the present thesis contributes to the new development and application of artificial neural networks in the workflow of medical imaging interpretation. Specifically, it introduces deep learning-based novel techniques to the fields of image reconstruction, image analysis, and privacy preservation. The current research aimed to study, implement, test, and validate the usage of deep neural networks in delivering a fully end-to-end learning-based medical data analysis.

In summary, the following main objectives can be identified:

- Development, implementation, and testing of a learning-based algorithm to enable fast high-resolution computed tomography image reconstruction from low-dose measurements;
- Development of a clinically-realistic evaluation tool for establishing the quantitative understanding of reconstruction quality;

- Development, implementation, and testing of a deep learning-based computer-aided diagnosis system that highlights suspicious findings in breast images;
- Development of a validation methodology for correctness, robustness and performance evaluation of the computer-aided diagnosis system;
- Extending existing methodology to two-view breast imaging analysis;
- Development of a secure noise-free homomorphic encryption scheme;
- Development of a generic deep learning library that exploits the homomorphic cryptography as a mechanism for enabling computations on sensitive data without disclosing patient-related health information;
- Development of a methodology for evaluating the performance of the deep learning library on homomorphically encrypted data;
- Development, implementation, and evaluation of a secure and privacy-preserving deep learning-based X-ray coronary angiography medical images analysis;
- Development, implementation, and evaluation of a secure and privacy-preserving deep learning-based whole-body circulation hemodynamic analysis;
- Identifying and examining both the level of security and utility of the encryption cryptosystem;
- Strengthening the encryption scheme security, while maintaining the performance and potential to be used in real-world applications;
- Development and implementation of a secure and privacy-preserving health risk score prediction.

1.3 Thesis Structure and Content

The thesis is organized as follows:

In **Chapter 2** a general introduction into the field of deep learning is provided. The chapter covers the basics of artificial neural networks. Moreover, the current trends in terms of network architectures are described in details, and the main tasks that can be tackled by learning-based models are captured, all linked by the common thread of using deep neural networks in the healthcare industry.

In **Chapter 3** a novel end-to-end deep learning-based framework is proposed for solving inverse problems in medical imaging. The framework is designed to encapsulate the knowledge of the physical model of computed tomography (CT) image formation and to produce high-quality images that account for human perception through a generative adversarial network with Wasserstein distance and a contextual loss.

In **Chapter 4** a framework is proposed to tackle the breast mass detection in digital breast tomosynthesis images. The solution consists of two modules, deep learning-based mass detection, i.e., identifying the locations of candidate lesions, and mass matching in ipsilateral tomosynthesis views. To improve the detection robustness, a registration of suspicious candidates provided by the learning-based model is employed upon the bilateral craniocaudal (CC) and mediolateral oblique (MLO) views of a breast. The algorithm mimics the radiologist's image interpretation routine using the basic notions behind the breast image formation.

Chapter 5 focuses on designing fully automated data-driven personalized-based medicine solutions by protecting the integrity of patient health data. A symmetric-key fully homomorphic encryption scheme is introduced as a potential solution for privacy-preserving computations within deep learning models. The applicability of incorporating homomorphic encryption into deep learning is showcased by tackling three different problems: digit recognition, whole body hemodynamic

analysis, and coronary angiography view classification. For each application, both the training and the inference phase are addressed and show that both can be performed on homomorphically encrypted data.

Finally, **Chapter 6** draws the final conclusions and summarizes the major findings. Additionally, it presents the personal contributions, the dissemination of the results, as well as give insights into the future directions of research.

Portions of this thesis were previously published as part of [3–6].

2. Deep Learning in Medicine

Since their first appearance in 1943 [7], the functionality of neural networks has continuously been associated with the way people learn and process information. More specifically, they were designed to emulate the synaptic connections between brain neurons, and later on, became the foundation of deep learning.

2.1 Neural Networks

On a high level, a neural network can be defined as a computational model that maps inputs to outputs through a composition of layers with interconnected processing blocks (transformations and activation functions). The architecture of a simple neural network is depicted in Figure 2.1. To allow for a complex arbitrary functional mapping, non-linear activation functions are typically added at each processing block. They filter the information that passes through the network, determining what input signal is relevant to be forwarded to the following layer. Mostly they decide whether a certain neuron should be activated or not, and without them, the neural network becomes a simple linear model.

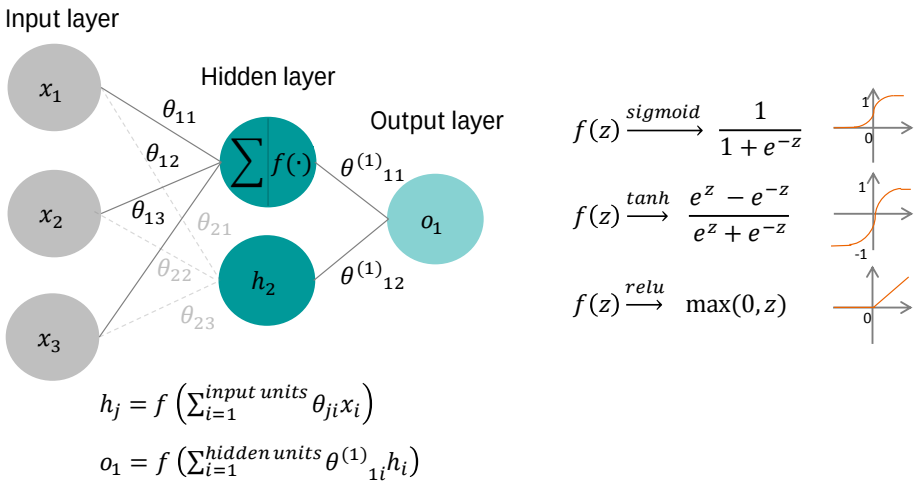


Figure 2.1: The architecture of a simple neural network described by the input, output, and a hidden layer in-between. Information flows through all layers, starting from the input layer to the output layer. Herein, every neuron receives information from all neurons of the previous layer (in literature called a fully connected layer). The connections θ between the processing blocks are the parameters that have to be adjusted in accordance with data and the formulated problem. Each processing block performs a transformation (herein a weighted sum of the input parameters), and the result is passed to an activation function f that will be used to add non-linear properties in the network. Activation functions are usually selected from a set of limited functions with certain mathematical properties. Non-linearity is needed to allow for a complex arbitrary functional mapping between input and output data.

Like any other machine learning methods, neural network models aim at learning from past experience to make predictions based on new observations. In supervised learning, during training phase the model automatically learns the mapping function (parameters of the model) based on labeled training examples in an iterative fashion by gradually making an adjustment.

Upon training, the network should be able to provide results that are statistically similar to the expected ones even when presented with input data never encountered by the network during training. Consequently, neural networks can be used in predicting an output from certain input features, classifying data, and even localizing patterns or objects in images.

2.2 Going Deeper: Deep Neural Networks

In essence, a deep neural network is nothing else than a neural network model composed of several layers of processing blocks and organized as an input layer, followed by multiple hidden layers and an output layer. Over the years, it has been shown that such an architecture facilitates the modeling of highly complex functions, allowing for the learning of richer intermediate representations. Hence, the key difference between shallow and deep neural networks is given by the depth of the models, although not standardized, typically a network with depth higher than two falls into the deep learning category.

2.3 Deep Learning in Medical Imaging

The current trend in medicine is towards tailored diagnosis, treatment planning, and disease prevention by identifying and correlating massive amounts of patient data (e.g., symptoms, diagnosis, treatments, etc.). In medical imaging analysis, deep learning found its way into data-driven medicine as a way of automatically understanding the semantic content of patient images for the diagnosis, detection, and segmentation of anatomical structures or diseases. Consequently, in recent years, DL has offered many data-driven solutions designed to tackle different imaging tasks (Figure 2.2), including segmentation, object detection, classification and, more recently, image generation (e.g., enhancement, denoising, reconstruction, registration etc.).

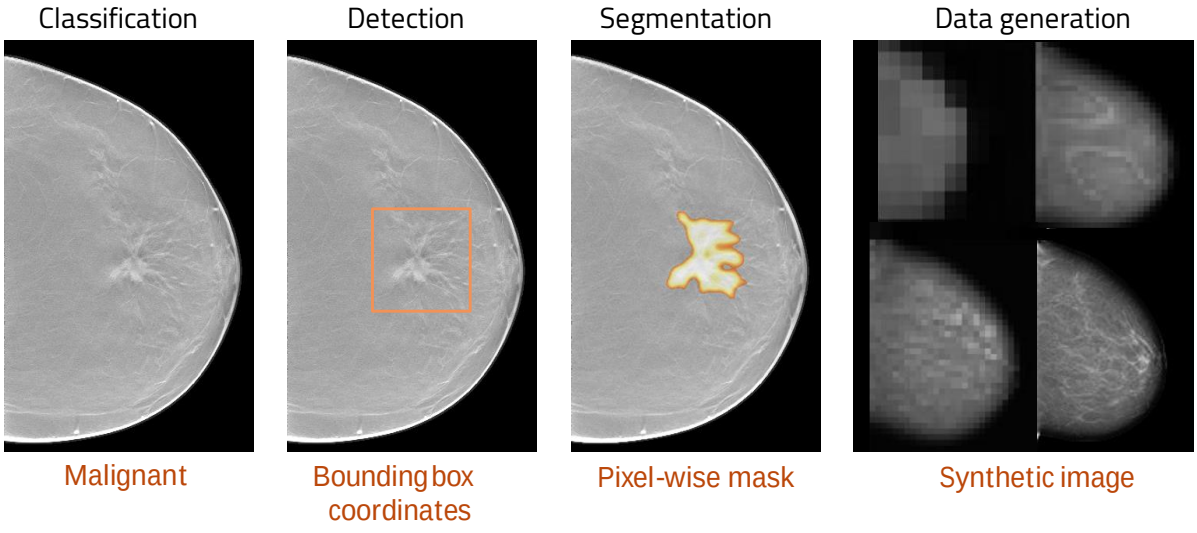


Figure 2.2: An overview of imaging-related deep learning tasks.

3. Towards Data-Driven CT Imaging Reconstruction

3.1 Introduction

X-ray computed tomography (CT) is a standard imaging modality used in clinical routines to reveal the internal structures of the human body. There is a clear dependence between the quality of the tomographic reconstruction and the diagnostic accuracy [8]. One way for obtaining higher quality CT scans implies an increased radiation dose, which also increases the potential risk of the patient to develop radiation-related pathologies. Meanwhile, by decreasing the dose, artifacts and noise are enhanced in the CT scan, deteriorating image quality and, hence, compromising the diagnostic accuracy.

Traditionally, analytical and iterative methods are the main approaches for performing image processing, including denoising, inpainting, restoration, and superresolution. The deep learning-based solutions for image reconstruction generally fall into two categories: (i) image-to-image domain reconstruction, and (ii) data-to-image domain reconstruction. Methods in the former category formulate the inverse problem as an image domain denoising task: this is one of the most straightforward ways of incorporating the learning process into the reconstruction. Usually, the reconstruction consists of two cascaded operations: an initial CT scan is obtained by applying a simple, direct inverse operator such as filtered back-projection, which is then post-processed by a Convolutional Neural Network (CNN) to reduce artifacts and noise.

Despite the impressive denoising results obtained in image-to-image reconstruction, details may be lost during the refinement step, as prior knowledge is usually neglected. Hence, data-to-image methods that incorporate prior knowledge on imaging physics have the potential of suppressing undesired effects, while preserving important details. Hence, more recently, the idea of learning a complete data-driven reconstruction has been investigated in several scientific works [9].

Therefore, methods in the second category aim at learning the entire reconstruction operation directly from the measured data, in an end-to-end manner. Herein a fully data-driven deep reconstruction model is proposed to address the low-dose CT reconstruction problem. While the Learned Primal-Dual algorithm proposed in [10] led to impressive results, the method suffered from a global over-smoothing effect, caused by the mean squared metric employed by the optimization procedure. To overcome this limitation, an adversarial training strategy is proposed to optimize the primal-dual reconstruction by also encouraging a human perceptual similarity, which enhances structural and textural details. Moreover, a patch-wise discriminator is adopted to further improve the authenticity of reconstructions: the model is constrained to gradually optimize the synthetic reconstruction image with emphasis on the region identified as being unrealistic.

3.2 Mathematical Formulation of CT imaging

X-ray computed tomography (CT) is an imaging procedure used to produce detailed 3D anatomical images. Typically, the process of generating a CT image includes two steps: data acquisition and image reconstruction. The CT imaging process is depicted in Figure 3.1.

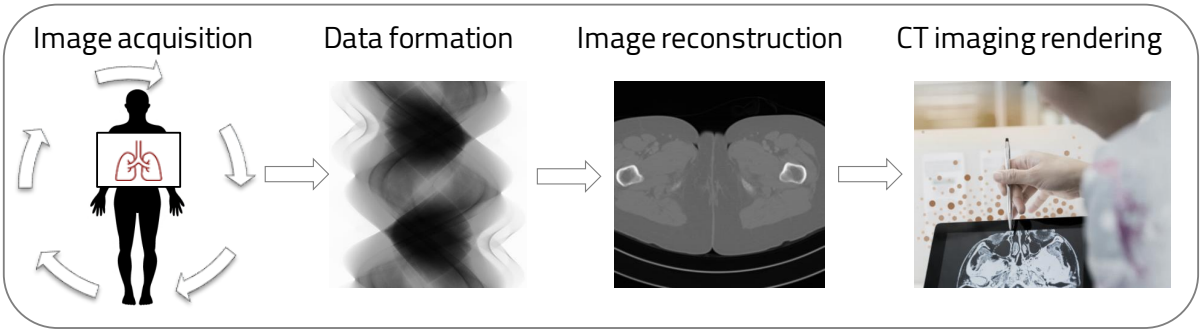


Figure 3.1: The process of CT imaging, from data acquisition to volume rendering.

During the acquisition, an object is placed between the X-ray source emitting a cone beam of X-rays, and a flat panel detector which captures the X-ray absorption of different tissues (materials) along a straight line path L. The retrieved information is known as measurement or projection data. To observe the object from different directions, the source-detector pair is rotated around the object, capturing hundreds of projections. The projections generated from equally spaced angular positions are stacked to form a so-called sinogram. The sinograms are then converted into tomographic 3D image slices during a reconstruction process. Finally, the generated slices are stacked together to form the 3D image of the patient, which allows for identification and localization of tissues, organs, abnormalities, etc.

Hence, mathematically, the goal of X-ray tomography can be formulated as recovering the attenuation map of the observed object $f \in \mathcal{X}$ from measurements $g \in \mathcal{Y}$, where $\mathcal{X}$ denotes the reconstruction space and $\mathcal{Y}$ represents the measurement space. CT imaging could be considered as a forward operator (Radon transform) that creates the measurements g from image f, as:

$$g = H(f), \tag{3.1}$$

where $H : \mathcal{X} \to \mathcal{Y}$ is the system matrix.

A typical approach for solving the reconstruction algorithm is based on direct inversion, described in CT imaging by the Radon transform adjoint operator, i.e., back-projection, $H^* : \mathcal{Y} \to \mathcal{X}$, as:

$$f = H^*(g). \tag{3.2}$$

The reconstruction obtained on the basis of the conventional back-projection operator suffers from strong blurriness.

More recently, model-based iterative reconstruction has emerged as an alternative solution. Starting from an empty image, and incorporating assumptions regarding data acquisition, the algorithm iteratively updates the reconstruction, using both the backward and forward projections, to minimize an error metric. Hence, X-ray tomography can be formulated as an optimization problem as follows [11]:

$$\arg\min_{f \in \mathcal{X}} \|(H(f) - g)\|_2 + \gamma R(f), \tag{3.3}$$

where the term on the left is the data fidelity term that measures the discrepancy between the original measured projection data g and the synthesized projection data $H(f)$. The term on the right $R : \mathcal{X} \to \mathbb{R}$ is a regularization term that incorporates prior knowledge of f, introduced to improve well posedness, with the hyper-parameter γ controlling the balance between prior knowledge and data fidelity.

The optimization problem formulated in 3.3 is usually addressed using gradient-based methods. However, the main drawback of an iterative reconstruction algorithm is the large computation time.

3.3 Data-Driven Image Reconstruction

Thus, learning-based iterative schemes represent a promising alternative for the computationally demanding classical approach. This data-driven method, relying on deep learning, aims at finding the parameter $\theta \in \mathbb{Z}$ of the parametric operator (pseudo-inverse) H_θ^+ based on a training dataset (pairs of ground truth images and their corresponding measurements) [12], by minimizing an error metric between the actual and estimated reconstructions, such that $H_\theta^+(g) \approx f$.

In the learning-based primal-dual reconstruction, an iterative scheme is unrolled into a neural network, capable of performing full-view reconstruction, starting from raw sinogram data, by alternatively updating both the data and the reconstruction.

The neural network combines the information provided by the forward and backward operators with data specific image filtering, obtained through the convolutional neural network blocks, to generate an intermediate reconstruction improved at each iteration. Encapsulation of the knowledge on the physics of image formation has multiple advantages: preservation of more details and reduction of number of unknowns (leading to a relaxation of the requirement for having a large training dataset).

Hence, an update function is learned for each iteration in both the image and the data domain:

$$g_i = G_{\theta_i}(g_{i-1}, H(f_{i-1}), g), \tag{3.4}$$

$$f_i = F_{\theta_i}(f_{i-1}, H^*(g_i)), \tag{3.5}$$

where G_{θ_i} and F_{θ_i} are convolutional neural networks (Figure 3.2), $i = \overline{1, N}$ represents the current iteration, with the learned pseudo-inverse being the output of the last iteration.

3.3.1 End-to-End Training

The above described problem is traditionally formulated as a regression problem [10], where the neural network is trained to provide reconstructed images that closely resemble the original images by minimizing a pixel-wise error metric [e.g., mean squared error (MSE) and mean absolute error (MAE)], between the actual and the generated reconstructions. Both MSE and MAE tend to be suboptimal for the image generation task, encouraging blurriness in images, and discouraging the generation of textures and structural details [13].

Hence, to generate the reconstructed image, an end-to-end network, Perceptual Primal-Dual Reconstruction Network (Perceptual PD-WGAN) that operates directly on sinogram data is proposed. To enable a higher level of details in the reconstructed images, the MSE is replaced by a perceptual (content) loss [14].

To enforce the generation of images statistically indistinguishable from actual reconstructions, a generative adversarial network is used. In a generative adversarial setup, two neural networks are coupled: the first network (generator) tries to generate high-quality images from sinogram data, the second network (discriminator) has to distinguish between actual image samples (from the real data distribution), and generated image samples. The two networks are trained jointly, following competing objectives.

For training stability, a Wasserstein generative adversarial network (WGAN) with gradient penalty [15] is adopted, whose overall goal is to minimize the Wasserstein distance between the true data distribution and the generated data distribution. Thus, the problem is formulated as follows: given a sample f from the reconstructed image distribution $\mathbb{P}_f$, and g from the sinogram data distribution $\mathbb{P}_g$, the generator G learns to map data from one distribution $\mathbb{P}_g$ to another $\mathbb{P}_f$, while the discriminator D estimates the probability of a generated sample to be part of the real distribution $\mathbb{P}_f$. In the end, the learning objective for WGAN can be formulated as: the generator G maximizes the probability of mistakes made by the discriminator (i.e., probability of successfully fooling the discriminator D):

$$\min_G \max_D L_{adversarial} = L_{critic} + \lambda L_{penalty}, \tag{3.6}$$

$$L_{critic} = \mathop{\mathbb{E}}_{\tilde{f}\sim\mathbb{P}_g}[D(\tilde{f})] - \mathop{\mathbb{E}}_{f\sim\mathbb{P}_f}[D(f)], \tag{3.7}$$

$$L_{penalty} = \mathop{\mathbb{E}}_{\hat{f}\sim\mathbb{P}_{\hat{f}}}[(\|\Delta_{\hat{f}} D(\hat{f})\|_2 - 1)^2], \tag{3.8}$$

where L_{critic} accounts for the Wasserstein distance estimation, and $L_{penalty}$ accounts for the gradient penalty term, whose contribution is being balanced by a weighted parameter λ. In above equations $\tilde{f} = G(g)$ and represents an image generated by generator G from measurements g, whereas $\hat{f}$ describes a sample drawn from a uniform distribution $\mathbb{P}_{\hat{f}}$ (sampling is performed along straight lines between pairs of samples from $\mathbb{P}_f$ and $\mathbb{P}_g$).

Therefore, the objective of the generator is to make the discriminator believe that a generated sample is real, thus to minimize:

$$L_{generator} = - \mathop{\mathbb{E}}_{\tilde{f}\sim\mathbb{P}_g}[D(\tilde{f})]. \tag{3.9}$$

The objective of the discriminator is to better distinguish between real and generated samples, thus to maximize the probability of identifying the data:

$$L_{discriminator} = \mathop{\mathbb{E}}_{\tilde{f}\sim\mathbb{P}_g}[D(\tilde{f})] - \mathop{\mathbb{E}}_{f\sim\mathbb{P}_f}[D(f)] + \mathop{\mathbb{E}}_{\hat{f}\sim\mathbb{P}_{\hat{f}}}[(\|\Delta_{\hat{f}} D(\hat{f})\|_2 - 1)^2]. \tag{3.10}$$

While the original WGAN uses a global discriminator, herein a patch-wise discriminator [16] is employed, which analyzes individual regions from the generated image and predicts a discriminative score map. The discriminative score provides constructive revisions for the generator, enforcing additional constraints for the network, leading to higher quality synthesized reconstructions.

Additionally, to encourage the generated images to appear more realistic, a perceptual loss is defined for the generator, based on the similarity of two images (the distance between high-level image features extracted from a pre-trained network).

$$L_{content} = \mathop{\mathbb{E}}_{g\sim\mathbb{P}_g, x\sim\mathbb{P}_f}[\|(\Phi(G(g)) - \Phi(f))\|_2^2], \tag{3.11}$$

where Φ is a feature map extracted by the network, $G(g)$ represents the generated reconstruction, and f is the true tomographic image. The last convolutional layer from the VGG19 network [17] is employed as a feature extractor.

Given the adversarial loss $L_{adversarial}$, and the content loss $L_{content}$, a final loss function is defined as a weighted sum of all individual losses:

$$\min_G \max_D L_{content} + \lambda_1 L_{adversarial} + \lambda_2 L_{supervision}, \tag{3.12}$$

Note that a supervision loss for a direct comparison of images in the reconstruction domain can also be incorporated as an additional constraint:

$$L_{supervision} = \mathop{\mathbb{E}}_{g\sim\mathbb{P}_g, f\sim\mathbb{P}_f}[\|G(g) - f\|_1], \tag{3.13}$$

The training procedure for the proposed PD-WGAN network is outlined in Algorithm 3.1.

3.3.2 Network Architecture

The architecture of the proposed Perceptual Primal-Dual Reconstruction Network (Perceptual PD-WGAN), that addresses the challenging sinogram-based iterative tomographic reconstruction problem, is depicted in Figure 3.2.

Algorithm 3.1 PD-WGAN training algorithm

Input: Generator G with parameters θ, Discriminator D with parameters ω, the pre-trained VGG-19 network.

Input: The forward and adjoint operators H, H^*, the number of iterations for the reconstruction N, the primal and dual spaces N_{primal}, N_{dual}.

Input: The training set $\{\boldsymbol{g}^{(j)}, \boldsymbol{f}^{(j)}\}_{j=1}^{M}$, the number of training epochs N_{epochs}, the batch size b, the number of iterations for the discriminator and generator N_D, N_G, the weighting factors $\lambda, \lambda_1, \lambda_2$, Adam hyperparameters α, β_1, β_2.

Output: Updated parameters θ and ω.

procedure Train($\boldsymbol{g}, \boldsymbol{f}$)
  Initialize network parameters: θ, ω
  for $epoch = 1, \ldots, N_{epochs}$ **do**
    for $critic = 1, \ldots, N_D$ **do**
      Sample a batch of b training samples: $\{\boldsymbol{g}^{(i)}, \boldsymbol{f}^{(i)}\}_{i=1}^{b} \sim \{\boldsymbol{g}^{(j)}, \boldsymbol{f}^{(j)}\}_{j=1}^{M}$
      for $i = 1, \ldots, m$ **do**
        Sample a random number: $\epsilon \sim Uniform[0, 1]$
        Generate reconstructed image: $\tilde{\boldsymbol{f}}^{(i)} \leftarrow G(\boldsymbol{g}^{(i)}, H, H^*, N, N_{primal}, N_{dual})$
        Compute: $\hat{\boldsymbol{f}}^{(i)} \leftarrow \epsilon \boldsymbol{f}^{(i)} + (1 - \epsilon)\tilde{\boldsymbol{f}}^{i}$
        Compute loss: $L_D^{(i)} \leftarrow D(\tilde{\boldsymbol{f}}^{(\boldsymbol{i})}) - D(\boldsymbol{f}^{(\boldsymbol{i})}) + \lambda(\|\Delta \hat{\boldsymbol{f}}^{(\boldsymbol{i})} D(\hat{\boldsymbol{f}}^{(\boldsymbol{i})})\|_2 - 1)^2$
      end for
    end for
    Update the Discriminator D: $\omega \leftarrow$ Adam$(\omega, L_D, \alpha, \beta_1, \beta_2)$
    for $generator = 1, \ldots, N_G$ **do**
      Sample a batch of b training samples: $\{\boldsymbol{g}^{(i)}, \boldsymbol{f}^{(i)}\}_{i=1}^{b} \sim \{\boldsymbol{g}^{(j)}, \boldsymbol{f}^{(j)}\}_{j=1}^{M}$
      for $i = 1, \ldots, m$ **do**
        Generate reconstructed image: $\tilde{\boldsymbol{f}}^{(i)} \leftarrow G(\boldsymbol{g}^{(i)}, H, H^*, N, N_{primal}, N_{dual})$
        Compute content loss: $L_{content}^{(i)} \leftarrow \|VGG(\tilde{\boldsymbol{f}}^{(\boldsymbol{i})}) - VGG(\boldsymbol{f}^{(\boldsymbol{i})})\|_2^2$
        Compute supervision loss: $L_{supervision}^{i} \leftarrow \|G(\tilde{\boldsymbol{g}}^{(\boldsymbol{i})}) - \boldsymbol{f}^{\boldsymbol{i}}\|_1$
        Compute loss: $L_G^{(i)} \leftarrow \lambda_1 L_{content}^{i} + \lambda_2 L_{supervision}^{(i)} + D(\tilde{\boldsymbol{f}}^{(\boldsymbol{i})})$
      end for
    end for
    Update the generator G: $\theta \leftarrow$ Adam$(\theta, L_G, \alpha, \beta_1, \beta_2)$
  end for
end procedure

3.4 Experiments

To benchmark and evaluate the proposed algorithm, a realistic clinical application was considered: reconstruction of full-dose computed tomography from simulated low-dose projection data (sinograms). The training samples contain abdominal CT scans from a dataset made available to participants in an NIH , AAPM and Mayo Clinic sponsored Low Dose CT Grand Challenge [18]. The dataset includes volumes with 3 mm and 1 mm slice thickness corresponding to full-dose abdominal CT scans from 10 patients. For the proposed experiments, samples from the 3 mm slice thickness reconstructions were extracted and the dataset was split at patient level into 2198 512×512 CT slices for training, and 210 CT slices for validation.

As the proposed network incorporates knowledge on the geometry of the CT system in the form of forward and backward operators, an imaging model was defined to synthetically generate low-dose projection data starting from the full-dose reconstructions.

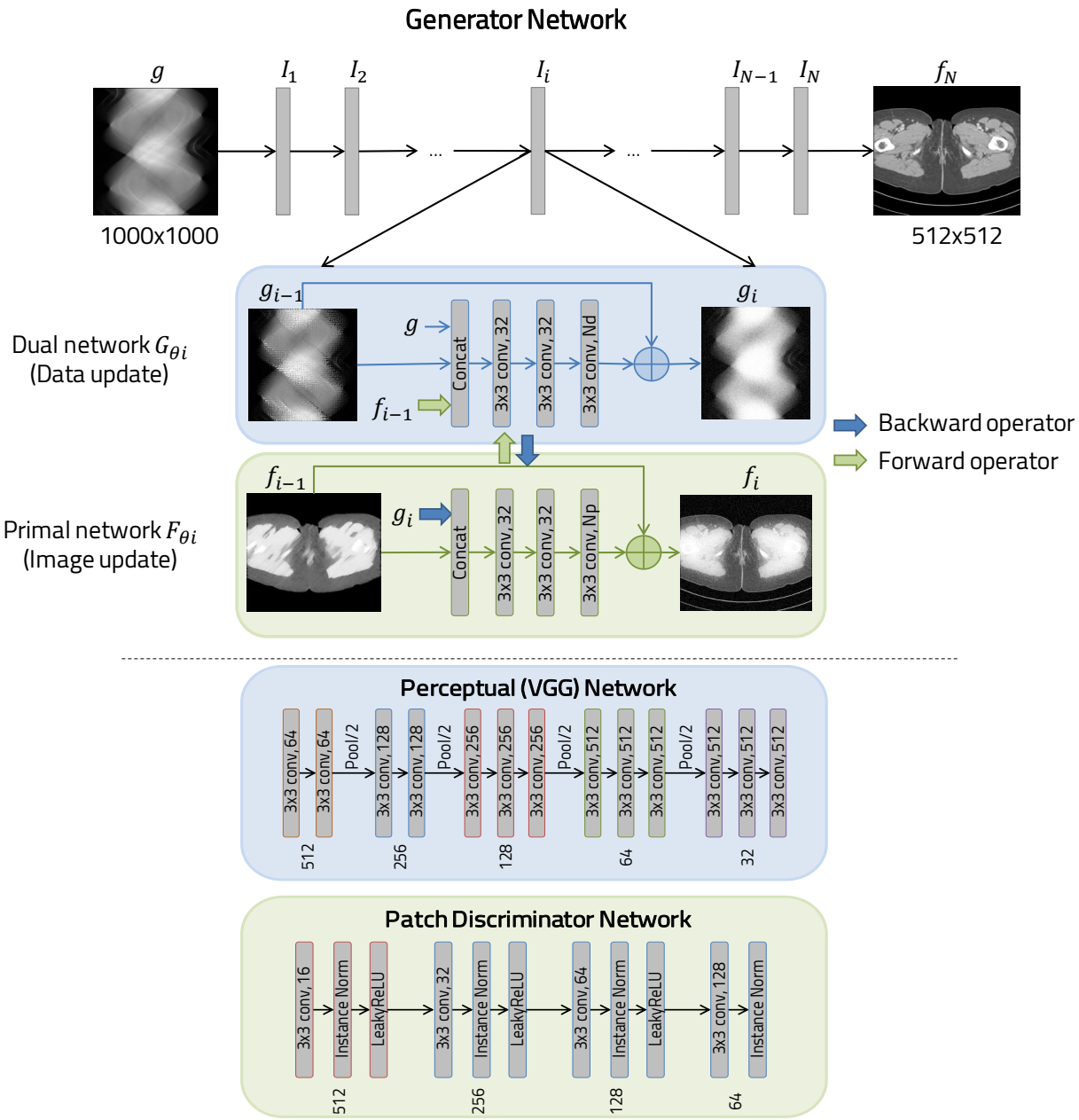
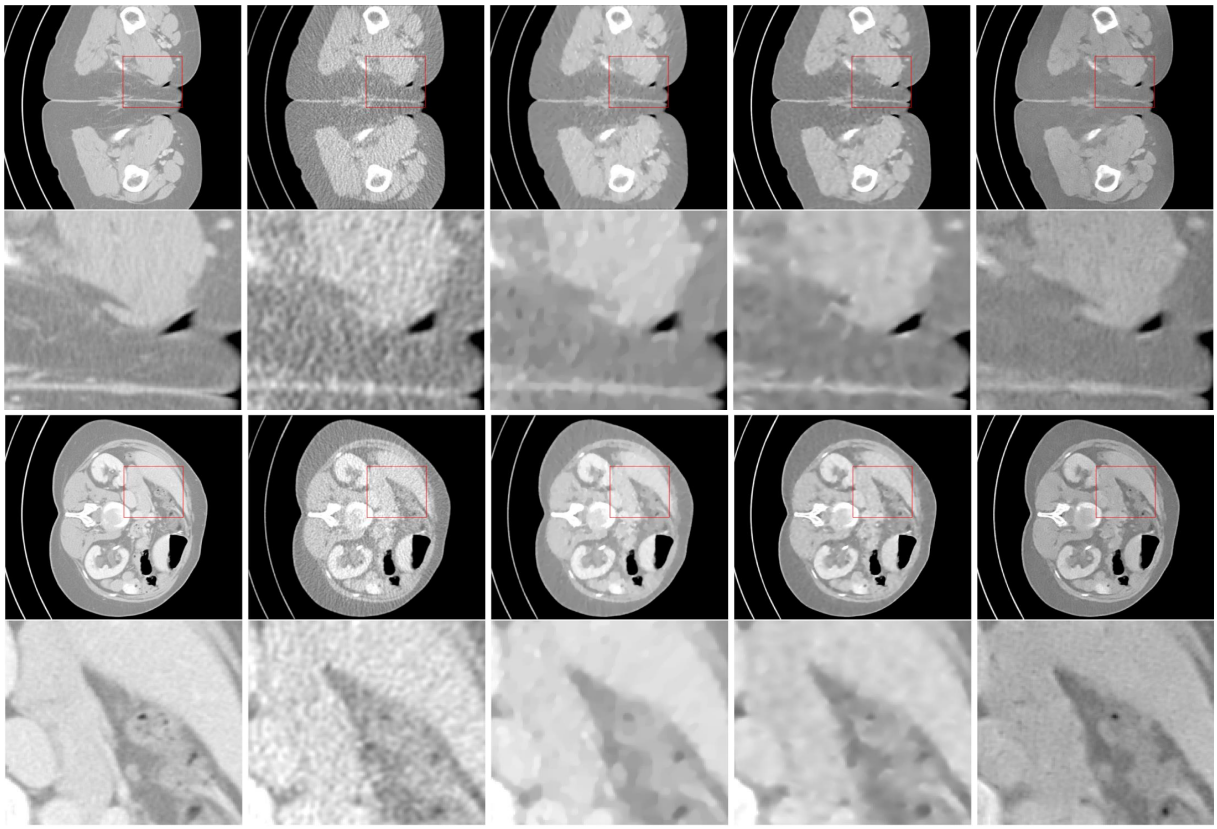


Figure 3.2: Overall structure of the proposed Perceptual PD-WGAN network for solving a tomographic problem. The iterative scheme is unrolled into $N = 8$ iterations with independently trainable parameters and 5 filter maps are generated in both the primal and dual path $Nd = Np = 5$ for information persistence between iterations.

3.5 Results

The Learned Primal-Dual network trained with mean squared error (PD-MSE) was implemented as the main reference method of data-driven unrolled networks. The structure of the network was identical to the one in the original paper [10]. Henceforth, the proposed solution (Perceptual PD-WGAN) was compared against three reconstruction algorithms, including the classical Filtered Backprojection (FBP) with Hann window, the TV regularized reconstruction, and the state-of-the-art Learned Primal-Dual (PD-MSE).



a) Full Dose GT b) FBP c) TV d) PD-MSE e) Perceptual PD-WGAN

Figure 3.3: Full-dose tomographic reconstructions: abdominal cross-sectional CT images alongside a zoomed in region delimited by the red rectangle. The display window is [-500, 200]HU.

3.5.1 Qualitative Evaluation

To assess the quality of the generated reconstructed tomographic images, representative slices from the testing dataset were selected, and results are depicted in Figure 3.3. As shown in Figure 3.3, the Perceptual PD-WGAN can mitigate the over-smoothing effect to a certain extent. However, compared to the reference image, there are still some details that appear to be noisy, and structures that tend to be slightly distorted.

To assess the reconstruction quality, with emphasis on the clinical use, a blinded randomized reading was performed by three independent readers, from which two were experienced radiologists. The image quality was assessed on the five-point Likert scale from three perspectives: noise, artifacts and overall diagnosability.

The computed mean of readers' quality ratings alongside the individual scores are shown in Figure 3.4. On the five-point comparative scale, the proposed PD-WGAN reconstruction method achieved the highest scores on all three criteria. All reviewers agreed that PD-WGAN based reconstructed images have less noise and artifacts than was perceived for the PD-MSE images. Moreover, in agreement with noise and artifact suppression, readers identified PD-WGAN reconstruction method as delivering higher quality images for diagnostic acceptability.

3.5.2 Quantitative Evaluation

For a quantitative assessment of reconstruction quality, Table 3.1 displays the mean values and the standard deviations for two typically used image quality metrics: Peak Signal to Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM). Additionally, the reconstruction runtime for

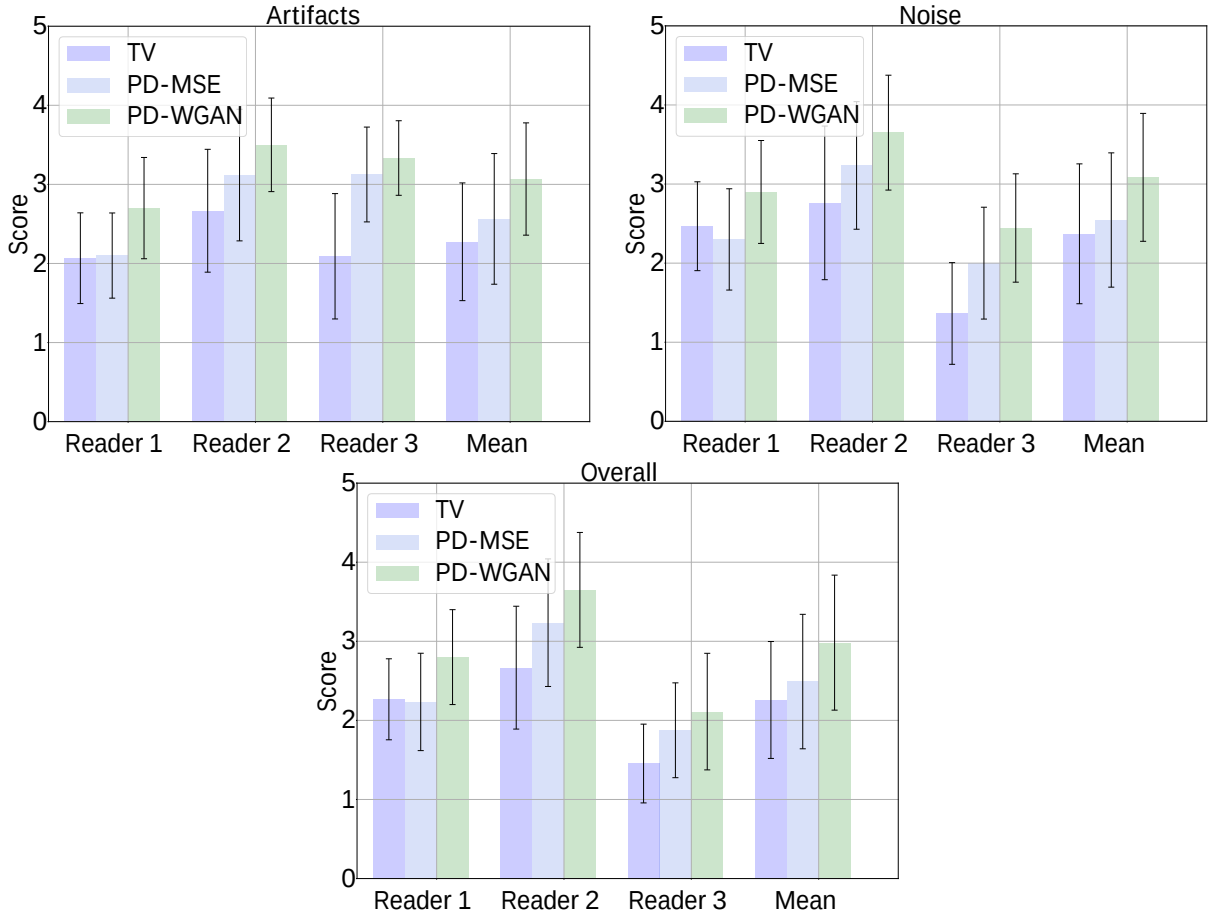


Figure 3.4: Mean qualitative scores for the three readers on a five-point Likert scale (1 = unnacceptable, 2 = poor, 3 = acceptable, 4 = good, and 5 = excelent). The mean scores with regard to the overall image quality for diagnosis, image noise and artifact supression is shown for each reader individually. Lastly, the overall averaged scores is also depicted.

each of the considered algorithms is also displayed in Table 3.1.

Table 3.1: Quantitative comparison of different reconstruction algorithms.

Method	Evaluation Metrics		Runtime (s)
	SSIM	PSNR	
FBP	0.827±0.018	34.533±1.451	0.452±0.084
TV	0.940±0.002	37.887±1.432	59.581±0.829
PD-MSE	0.954±0.004	42.598±1.312	0.537±0.137
Perceptual PD-WGAN	0.911±0.006	31.375±2.218	0.520±0.176

Although the results obtained by the proposed model are perceptually more similar to the reference images, this is not reflected by the PSNR and SSIM metrics. However, this is not surprising, as the proposed tomographic network was trained to minimize a feature-based loss in an adversarial manner, not with the intention of reducing pixel-wise differences, but rather for encouraging the generation of more realistic images.

While training the Perceptual PD-WGAN is more time consuming, when compared to the prior Learned Primal-Dual method, the reconstruction runtime is approximately the same, as the adversarial structure of the network is neglected during inference. Furthermore, the data-driven recon-

struction is much faster, i.e., two orders of magnitude, than the classical TV regularized iterative method.

3.6 Discussions and Conclusion

It has been shown in a clinically realistic scenario that a direct inversion operator can be learned from pairs of measurement-reconstruction samples, without relying on any inversion initialization. As opposed to denoising deep learning-based post-processing methods, where the inverse problem is treated as an image-to-image method, the proposed model relies on physical and mathematical knowledge, and, by resembling the iterative optimization problem, it offers a combination of tracing and understanding the network involvement in image reconstruction. Although promising results have been reported in literature for learning-based post-processing methods, they are limited by the information available during the training phase. As most of them are initialized with filtered back-projection, for low-dose and sparse sampling problems, the information is lost during the initial reconstruction and cannot be restored. Moreover, compared to post-processing methods, the proposed framework not only embeds prior useful knowledge, but it also allows for the correction to be performed in both the image and the projection domain.

The proposed method has been evaluated on a low-dose inverse problem, using the Mayo clinic CT dataset, and obtained promising tomographic images, to be further evaluated for diagnosis, as compared to the state-of-the-art MSE based Learned Primal-Dual reconstruction network. A runtime feasible for a routine clinical setting was maintained. In the proposed experiments, Perceptual PD-WGAN model led to encouraging qualitative results in terms of noise suppression and texture preservation, but there is still room for improvement in detail enhancement. The results prove that the advanced loss function employed in the current work represents a better alternative for the classical MSE-based optimization, known to produce sub-optimal qualitative results, due to its over-smoothing tendency. The conclusion with regard to the reconstruction quality have been drawn based upon the qualitative and quantitative measurements. Although PSNR and SSIM are no longer regarded as reliable indicators of image quality, they are still used as a standard evaluation approach. For a more in-depth validation of the image quality for clinical use, blind readings were performed by expert radiologists. The results have indicated that the proposed Perceptual PD-WGAN reconstruction model can increase the perceived quality of the image with regards to overall diagnosability, artifact and noise suppression, as evaluated by expert radiologist, which supports the use of the algorithm for improved data-driven low-dose CT reconstruction.

In conclusion, an end-to-end deep learning based framework is proposed for solving inverse problems in biomedical imaging, yielding promising results for full-dose tomography reconstruction, starting from low-dose measurements. The proposed deep learning-based solution relies on the idea of unfolding an iterative image reconstruction algorithm into finite iterations represented by a deep learning network, in which a primal-dual optimization is interpreted as the generator block, within a Wasserstein Generative Adversarial Network architecture. Furthermore, the framework integrates prior knowledge information regarding the CT imaging formation, and improves the training strategy by imposing human image quality perception. The algorithm qualitatively outperforms the previously proposed approach for data-driven deep learning based on primal-dual unrolled optimization, and the results provide promising evidence for its performance in inverse problems.

4. Towards Computer-aided Detection System in Digital Breast Tomosynthesis

4.1 Introduction

Although breast cancer is known to be one of the leading causes of cancer death among women, timely diagnosis and treatment can drastically reduce mortality [19]. Currently, the primary approach to assess the early signs of breast cancers is through X-ray mammography. While the use of mammography has shown to significantly reduce the mortality rate, it suffers from high recall rates and misclassifications [20].

Lately, a newer breast imaging technique, Digital Breast Tomosynthesis (DBT), with the potential of improving both sensitivity and specificity has rapidly emerged in the field [21]. DBT addresses the well-known limitations of 2D digital mammography by allowing a volumetric rendering of the breast, and thus reducing the tissue overlapping effects.

Typically, two standard DBT volumes are acquired of each breast corresponding to a bilateral craniocaudal (CC) and a mediolateral oblique (MLO) views. Hence, the breast is compressed from two non-orthogonal directions leading to additional useful information. Therefore, in a clinical routine, the use of a two-view analysis or sometimes even four-views, as shown in Figure 4.1, can improve the detection and diagnosis of abnormalities [22].

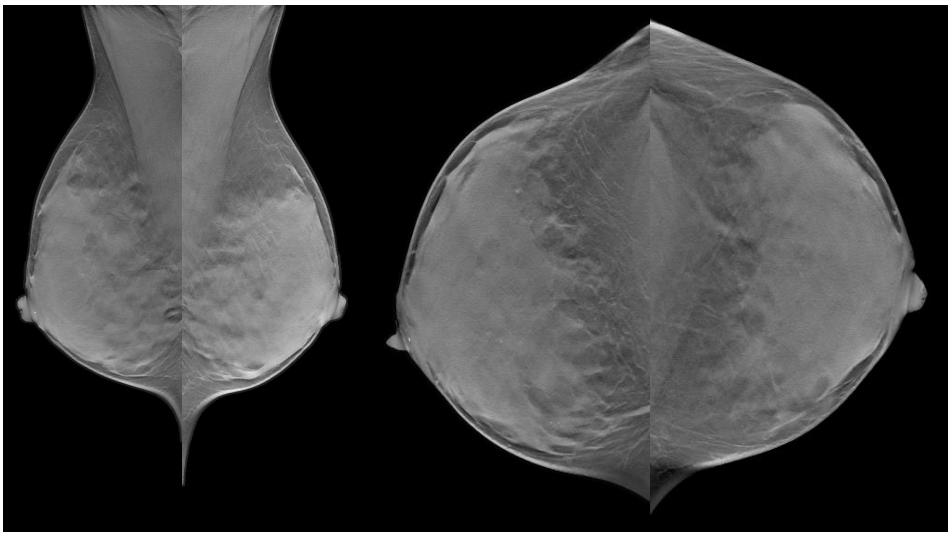


Figure 4.1: DBT slices of mediolateral oblique (left) and craniocaudal (right) views of a left and right breast of a patient.

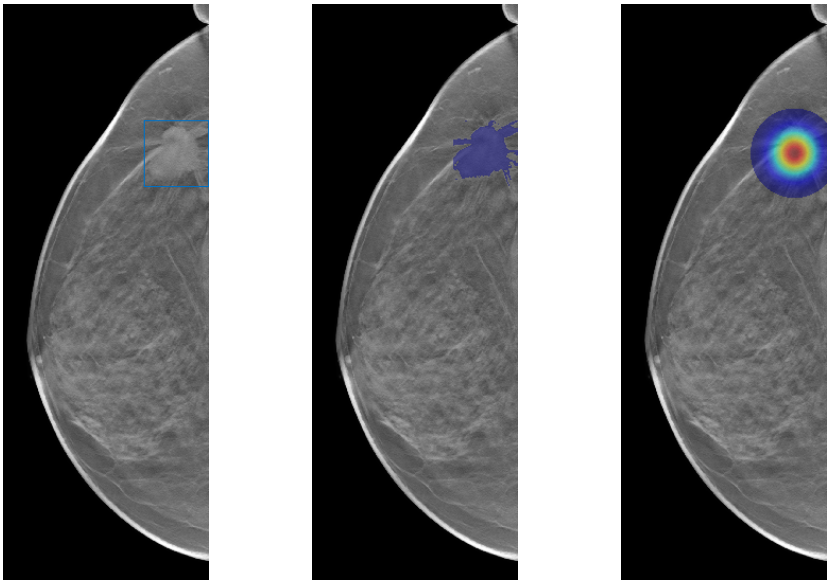
Even through DBT is a rising imaging technique due to its clear advantages in improving breast cancer assessment in clinical routines, mammography remains the gold standard screening technique. This is mainly because breast tomosynthesis is still a relatively new technique that comes at a higher acquisition cost and which, due to the larger amount of contextual information, requires additional time for interpretation.

Considering the above, herein a data-driven solution is proposed to tackle the mass detection in DBT breast volumes. The solution relies on an a deep fully convolutional neural network trained to localize masses in 2D DBT slices. To leverage the potential of data-driven models in breast imaging analysis, three main directions are followed: (i) reducing the DBT lack of data implication, (ii) addressing the problem of imprecise mass boundaries, and (iii) reducing the number of false-positive findings. A two-staged fine-tuning strategy is adopted to solve, to certain extent, the problems that arise due to the lack of an insufficiently large training DBT dataset. More specifically, the proposed solution takes into account the idea of features re-usability, where a series of layers are being initialized with weights that have been previously found by training the network on a different task on millions of images. The pre-trained model is first adapted to operate on mammography data and further fine-tuned on DBT data. Moreover, to address the problem of inaccurate transition between lesion and adjacent tissues, the detection task is formulated to facilitate the DBT mass identification and the training strategy alongside the loss function are adapted accordingly. Lastly, a two-view assessment framework is proposed to improve the recall rate by reducing the number of false suspicious findings through geometric mass mapping in DBT views.

4.2 Mass Detection

4.2.1 Problem Formulation

In medical imaging tasks oriented on object-of-interest identification aim at finding and highlighting regions where the objects of interest reside. When employing deep neural network in a supervised manner, the models learn to automatically extract such regions by minimizing an error between the manually annotated ground truth and the prediction.



(a) (b) (c)

Figure 4.2: DBT ground truth examples: slice showing a mass encoded with (a) box coordinates, (b) segmentation map, and (c) confidence map.

To reduce the mapping complexity, instead of aiming at directly regressing the coordinates of a bounding box (Figure 4.2a) or at classifying the image pixels (Figure 4.2b), the detection problem can be cast as a confidence map-based localization (Figure 4.2c). Hence, the ground truth is encoded as a 3D Gaussian heatmap (with same size as the observed image) centered at mass location as:

$$f(x,y,z) = e^{-\frac{(x-\mu_x)^2}{0.1\sigma_x^2}} \cdot e^{-\frac{(y-\mu_y)^2}{0.1\sigma_y^2}} \cdot e^{-\frac{(z-\mu_z)^2}{0.1\sigma_z^2}} \tag{4.1}$$

where x, y, z are 3D image coordinates, μ_x, μ_y, μ_z are the lesion center coordinates as given by the annotated 3D bounding box, and $\sigma_x, \sigma_y, \sigma_z$ are the width, height and depth of the box. Therefore, a neural network seeks to minimize a distance metric between the ground truth and predicted confidence map centered at lesion location.

By encoding the information at pixel level, the model can identify multiple objects in an image, with responses at different spatial locations, without requiring knowledge about the total number of expected objects. Additionally, as the heatmap provides lower confidence for pixels located at the object boundaries, it introduces a way of coping for inexact lesion borderlines.

4.3 Deep Learning-based Mass Detection

4.3.1 Clinical Dataset

In order to conduct learning-based experiments, an in-house DBT database is used and split at patient level into training, validation and test sets. The DBT image partitioning is performed such that all statistical properties are well preserved across the datasets.

Table 4.1 summarizes the DBT breast imaging datasets used in this study.

Table 4.1: DBT data splitting: training, validation and test sets.

Dataset	Category	No. of cases	No. of volumes	No. of slices	No. of unique VOIs	No. of VOIs	No. of unique ROIs
Train	Positive	271	345	5375	363	515 M	5785
	Negative	871	1318	74011	N/A	N/A	N/A
Validation	Positive	130	156	2500	176	337 M	2821
	Negative	384	531	29213	N/A	N/A	N/A
Test	Positive	172	228	3528	254	388 M	3740
	Negative	573	805	45241	N/A	N/A	N/A

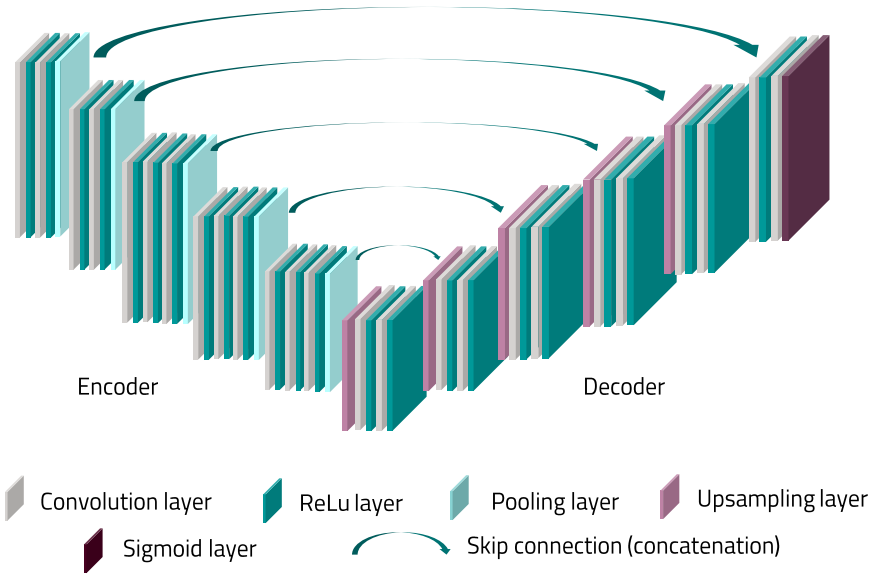


Figure 4.3: The architecture of the proposed learning-based mass detection model.

4.3.2 Network Architecture

Herein, an encoder-decoder architecture is proposed to tackle the mass detection problem in DBTs. The network (Figure 4.3) consists of an encoder i.e., contracting/downsampling path, responsible with the extraction of contextual features and a decoder, i.e., expanding/upsampling path, used to retrieve the image spatial information. Additionally, skip connections are added between the corresponding layers of the two paths to better recover the lost spatial information but also to capture the fine details. By doing so, the contextual information captured in the expanding path is combined with the location information that comes from the contracting path. Table 4.2 captures the topology of the proposed learning-based model.

Table 4.2: Configuration details of the proposed learning-based mass detection model.

Blocks	**Layers**	**# filters**	**patch size**
Input	-	-	(1024,1024,3)
DB+ MaxPooling	2 $\times$(3$\times$3 Conv-ReLU)	64	(1024,1024,64)
DB + MaxPooling	2 $\times$(3$\times$3 Conv-ReLU)	128	(512,512,128)
DB + MaxPooling	3 $\times$(3$\times$3 Conv-ReLU)	256	(256,256,256)
DB + MaxPooling	3 $\times$(3$\times$3 Conv-ReLU)	512	(128,128,512)
DB + MaxPooling	3 $\times$(3$\times$3 Conv-ReLU)	512	(64,64,512)
Upsampling + UB	3$\times$3 Conv-ReLU	512	(64,64,512)
	3$\times$3 Conv-ReLU	256	(64,64,256)
Upsampling + UB	3$\times$3 Conv-ReLU	512	(128,128,512)
	3$\times$3 Conv-ReLU	256	(128,128,256)
Upsampling + UB	3$\times$3 Conv-ReLU	512	(256,256,512)
	3$\times$3 Conv-ReLU	128	(256,256,128)
Upsampling + UB	3$\times$3 Conv-ReLU	256	(512,512,256)
	3$\times$3 Conv-ReLU	64	(512,512,64)
Upsampling + UB	3$\times$3 Conv-ReLU	128	(1024,1024,128)
	3$\times$3 Conv-ReLU	64	(1024,1024,64)
Output	3$\times$3 Conv-ReLU	32	(1024,1024,32)
	3$\times$3 Conv-sigmoid	1	(1024,1024,1)
# parameters	17M		

4.3.3 Network Training Details

Since the data-driven analysis on the complete 3D scan is still challenging to be performed as both the complexity of the network and the input dimension greatly impact the GPU memory usage, herein the focus lies on facilitating the data-driven CADe system development by tackling the mass detection problem through 2D data analysis. More specifically, the proposed mass detection model is trained to find the lesion centroid by regressing the heatmap. Due to the high in-plane spatial resolution, the full-size 2D DBT slices greatly limit the choice of network architecture and complexity. Thus, for more flexibility, training is performed on sub-regions, rather than on the full-resolution image.

To tackle the problem of having the number of pixels per structure, i.e., mass and background, not equally distributed inside the image, rather than finding the proper weights to weight the contribution of each structure to the final MSE loss function, the DBT mass detection network is optimized

by maximizing a continuous and differentiable variant of the Jaccard similarity coefficient [23, 24] between the target image T and the predicted probability map P:

$$L_{Jaccard} = 1 - \frac{\sum_i T_i \cdot P_i}{\sum_i T_i^2 + \sum_i P_i^2 - \sum_i T_i \cdot P_i} \tag{4.2}$$

It should be noted that only the positively labeled training observations contribute to the loss.

Herein a two-stage transfer of knowledge strategy is adopted, as shown in Figure 4.4. First, starting from a randomly initialized decoder and a pre-trained encoder (weights of the VGG19 [17] trained on ImageNet [25]), the network is adapted on 2D conventional mammograms to segment breast masses. Training is performed on 2D patches of 1024×1024 size. The Dice loss [26] is adopted as a segmentation loss function:

$$L_{Dice} = 1 - \frac{2\sum_i T_i \cdot P_i}{\sum_i T_i^2 + \sum_i P_i^2} \tag{4.3}$$

Thereafter, the resultant model is further fine-tuned on the final task on DBT data by minimizing the Jaccard loss.

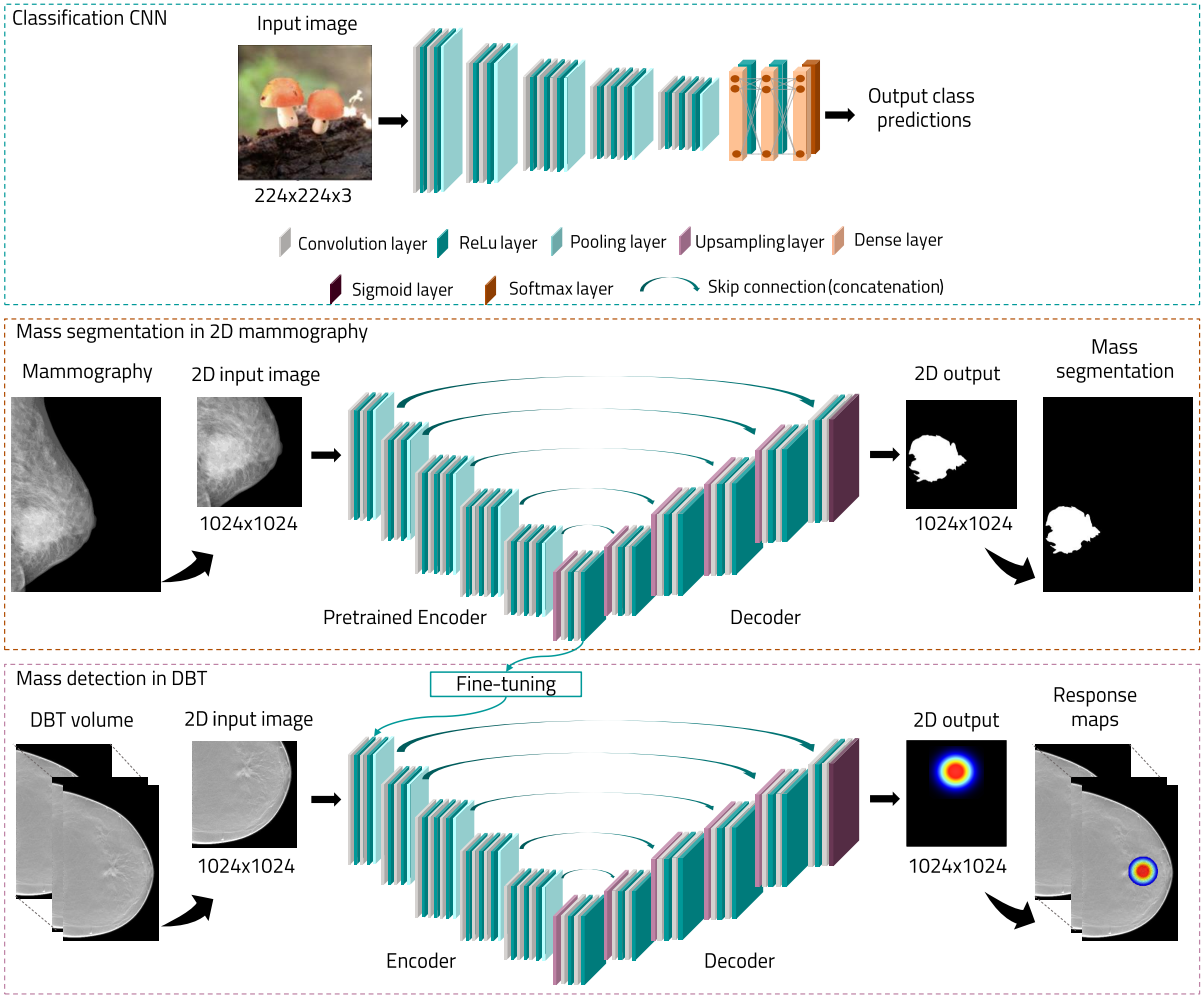
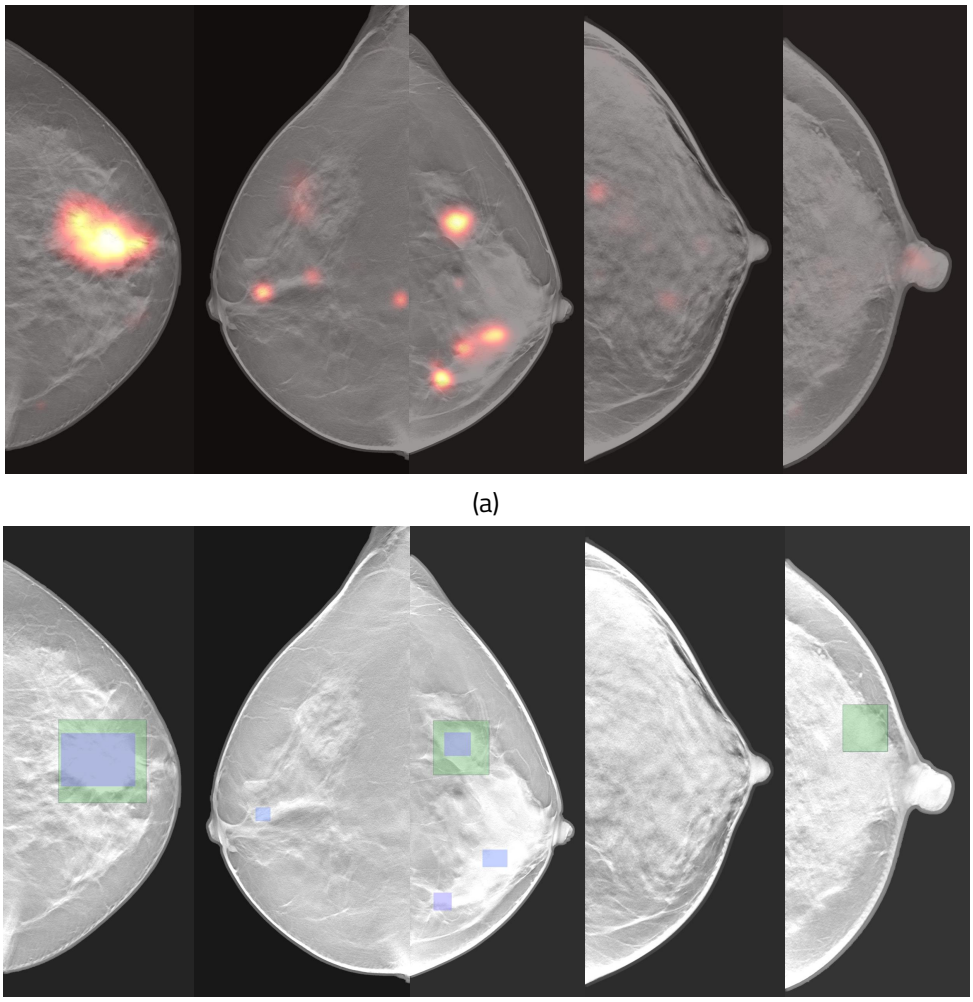


Figure 4.4: An overview of proposed learning-based mass detection training pipeline. The weights of convolutional layers of a classifier trained on ImageNet images are used to initialize the encoder of the proposed Unet-like models. Thereafter, the network is adapted to solve the mass segmentation in conventional 2D mammography images. Finally, the network is further fine-tuned for the mass detection problem on DBT images.

4.3.4 Results

Figure 4.5a exemplifies the network responses, i.e., output map, as obtained by the proposed model. To determine the corresponding final 3D lesion location, the network responses for each DBT slice are stacked and a mean-shift method is applied. In this stage not only the parameters of the bounding boxes are obtained but also a certainty measurement, i.e., the confidence score of the highlighted region as being a mass. Figure 4.5b shows examples of DBT mass detection results after the post-processing step. As shown in these images, the proposed data-driven mass detection model is able to identify masses of various sizes and appearances. However, there are also cases in which the model was unable to correctly identify the masses, especially when the mass appearance was hindered by dense breast tissues.



(a)

(b)

Figure 4.5: Examples of response maps as predicted by the mass detector model (a) and (b) the corresponding mass detection bounding boxes retrieved after the post-processing step. The boxes marked with green color represent the annotated bounding boxes and with blue are depicted the predicted locations. Each image describes the 3D DBT volume through the maximum intensity projection across the xy plane.

The performance of the proposed mass detection model is evaluated in terms of the Free Response Operating Characteristic (FROC) curve volume wisely. It measures the true-positive rate (TPR) against the number of false-positive findings per case, i.e., DBT volume. TPR is commonly refereed to as sensitivity or detection rate in clinical setting and describes the percentage of annotated masses that have been correctly identified by the model.

The obtained FROC curve on the test DBT dataset, for a detection threshold $T = 0.05$, is depicted in Figure 4.6. The plot indicates a detection rate of 80% at 0.7 FP on average on each DBT volume. However the balance between the two quantities can be controlled by altering the decision threshold. Herein the candidates are filtered based upon the confidence score of the highlighted region.

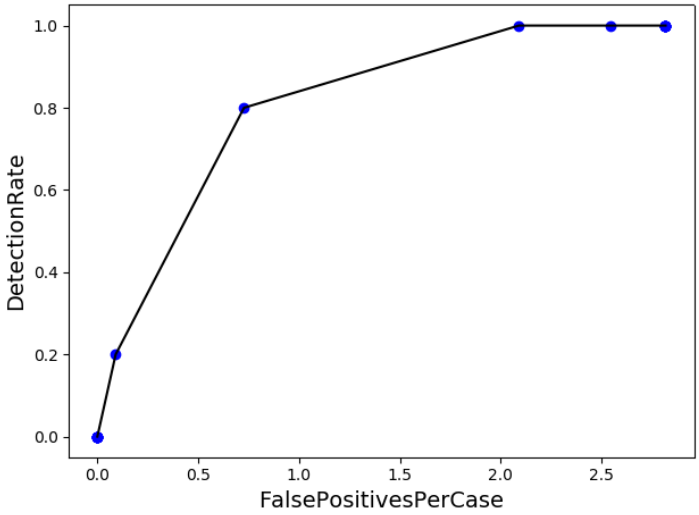


Figure 4.6: The Free Response Operating Characteristic (FROC) curve for the proposed DBT mass detection.

4.4 Mass Mapping in Ipsilateral Tomosynthesis Views

To make mass detection more reliable, radiologists combine the information acquired from CC and MLO views of a breast. A two-view assessment increases the chances of a lesion being seen in at least one of the views, but also improves the recall rate by eliminating false suspicious findings through view correlations [27]. However, due to compression and additional depth information, finding corresponding regions on different DBT views, in a clinical setup, is time consuming and prone to error. Moreover, there are cases where, due to the acquisition geometry, structures seen in a view may be only partially seen in the second view, as shown in Figure 4.7.

Although a two-view analysis of the same breast can increase the performance of a CADe system, as shown in 2D mammography studies [22], due to lack of data, most of the learning-based CADe systems for mass detection in DBT images rely on single view-information. An alternative solution that mitigates the data availability issue while maintaining the benefits of additional knowledge is through region correlation between the two views. Hence, instead of modeling a CADe system to detect a mass based on multiple input information, a mass location mapping stage can be added on top of a single-view CADe system to improve detection.

4.4.1 Geometric Mass Matching Criterion

Herein, a geometric matching solution that emulates the radiologist's technique is proposed. More specifically, the o'clock position of a finding is used to match corresponding regions in CC and MLO and improve lesion detection. By combining the information assumed from the geometry of the DBT with the nipple to lesion distance, an intermediary model can be assembled in the form of the o'clock system (Figure 4.8). Such a system may be used to map regions from one view to the other, without taking into consideration any compression mechanism or material properties, making the method fast and intuitive.

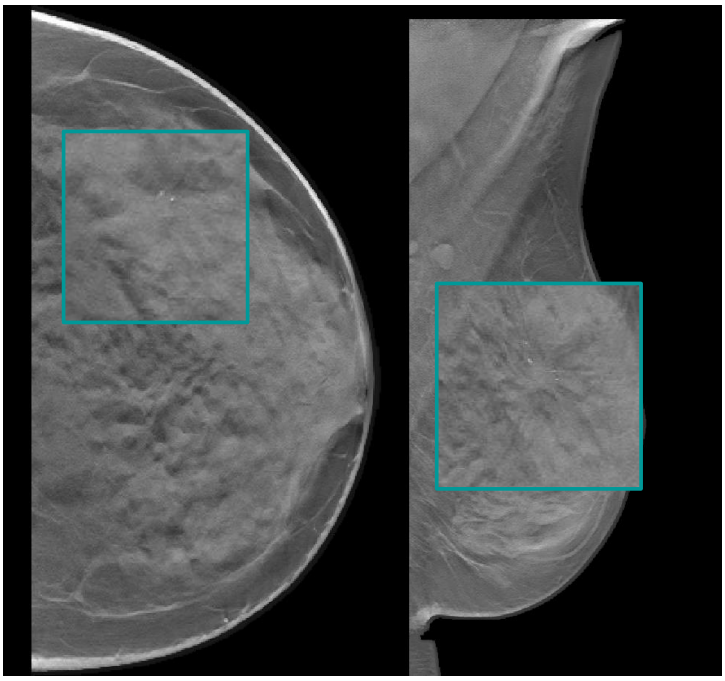


Figure 4.7: DBT mass example: a slice extracted from the CC and MLO tomosynthesis volumes that contains a mass depicted herein with the rectangular box. The mass is clearly visible in the MLO view (right) but only partially in the CC (left). The region is zoomed in for better visualization.

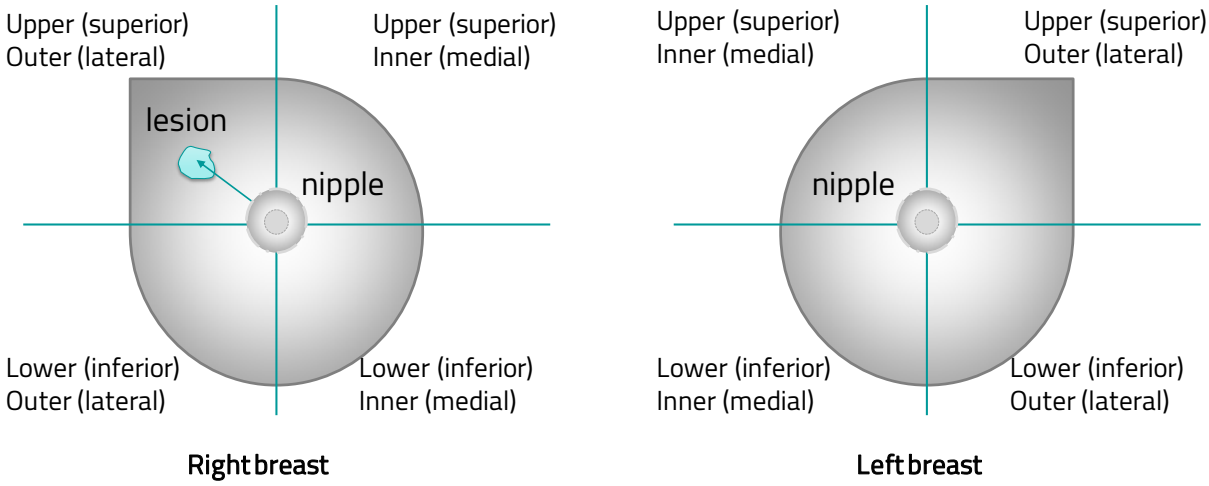


Figure 4.8: Quadrants in the o'clock system for the right and left breast.

Given the nature of the DBT breast imaging, the volumetric data provides an intuition about the position of the lesion in the breast. Therefore, in combination with the "scrolling" information, it can be established whether a lesion projected from the MLO view is median or lateral, or if it is superior or inferior for the CC view. For example, knowing that for an MLO view, X-rays pass from the upper inner towards the lower outer quadrant and based on the distance between the lesion position (i.e., the slice on which it is localized) and the edge of the breast, laterality can be infused. Thus, a 2-D slice of an MLO DBT view enables the first access to the lesion location in terms of superior/inferior direction, while by sliding through the volume, the location in the medio/lateral direction can be deduced. The same assumption is valid for CC DBT views but vice-versa. Figure 4.9 illustrates the proposed geometric pairing procedure.

Thus, given a 3D location in one view, the corresponding 3D location in the ipsilateral view is

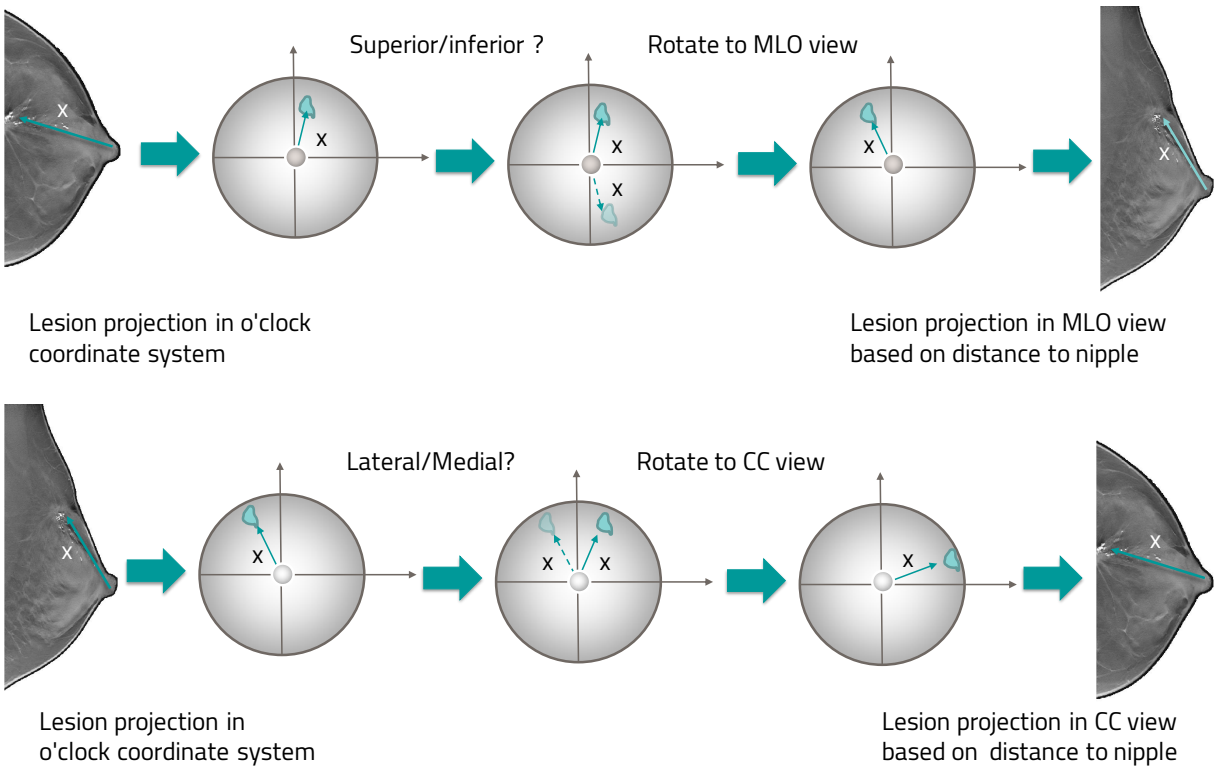


Figure 4.9: An overview of the in-plane (xy-plane) lesion location mapping algorithm from one DBT view to the other.

determined through a two-stage algorithm, as shown in Figure 4.10. Hence, for a given position in a view, the first stage of the algorithm provides the 2D corresponding position (in xy-plane) in the second view by projecting the information into the o'clock system and back to the target view (Figure 4.9). In order to find the correspondence between the 3D position of lesions on two DBT views, additional information related to the slice (z-axis) on which the lesion is located is required. In addition to the 2D information, the o'clock system gives also an intuition about the quadrant in which the lesion is located. Hence, in the second stage, the exact DBT slice is deduced by using a correlation metric between the sub-region extracted from the known lesion location in the input view and the potential regions from the target DBT volume. Therefore, based on the previously matched 2D location and the quadrant extracted from the lesion position in the o'clock system, a stack of possible candidates was formed by projecting the lesion location into DBT associated slices.

To measure the image similarity between paired sub-regions, a template matching approach was considered. The normalized cross-correlation between the lesion located in a view (I_s) and a paired region from the target candidate (I_t) stack was computed as follows:

$$R = \frac{\sum_{i,j}(I_s(i,j) - \bar{I_s})(I_t(i,j) - \bar{I_t})}{\sqrt{\sum_{i,j}(I_s(i,j) - \bar{I_s})^2}\sqrt{\sum_{i,j}(I_t(i,j) - \bar{I_t})^2}} \tag{4.4}$$

where $\bar{I_s} = \frac{1}{NM}\sum_{i,j} I_s(i,j))$ and $\bar{I_t} = \frac{1}{NM}\sum_{i,j} I_t(i,j))$ are the means of the patches. Herein, image I_s and I_t have the same dimension $N \times M$, with (i,j) denoting the index of a pixel in the patch.

The patches identified as having the maximum correlation value among all candidates govern the depth information required for the 3D lesion position mapping in DBT volumes.

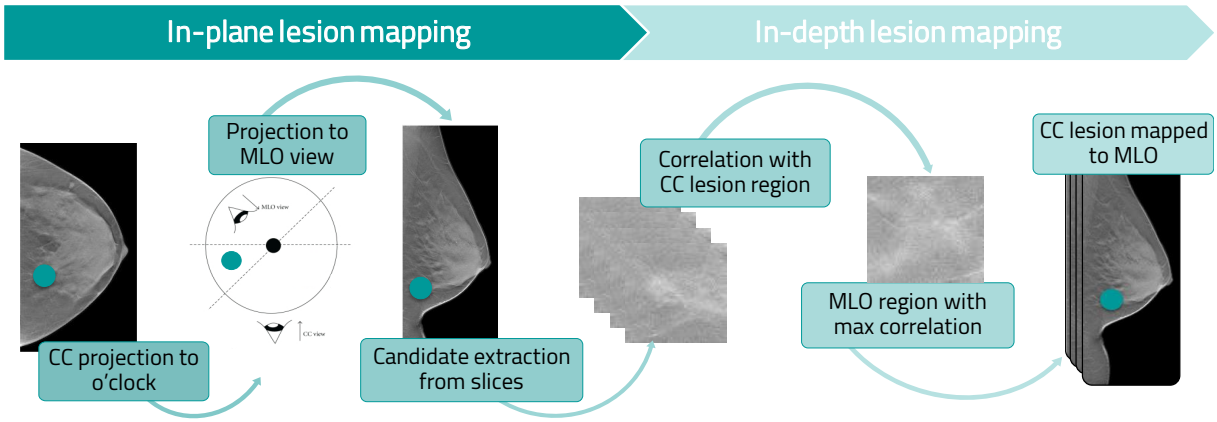


Figure 4.10: An overview of the proposed 3D lesion location matching in ipsilateral DBT views. The 3D coordinates of a CC lesion are identified in the MLO using the proposed two-stage algorithm. In the first step, the lesion is mapped in the in-plane space (xy-space) using a geometric matching criterion and the intermediary o'clock coordinate system. Next, the in-depth (z-axis) information is retrieved following a correlation metric criterion.

4.5 Results

The proposed framework for 3D position mapping in ipsilateral DBT views has been applied on DBT samples for which both CC and MLO volumes are available. The algorithm was first evaluated in terms of position correlation performances and secondly, the impact on joining the two-view information of the mass detection model was investigated.

4.5.1 DBT Position Correlation

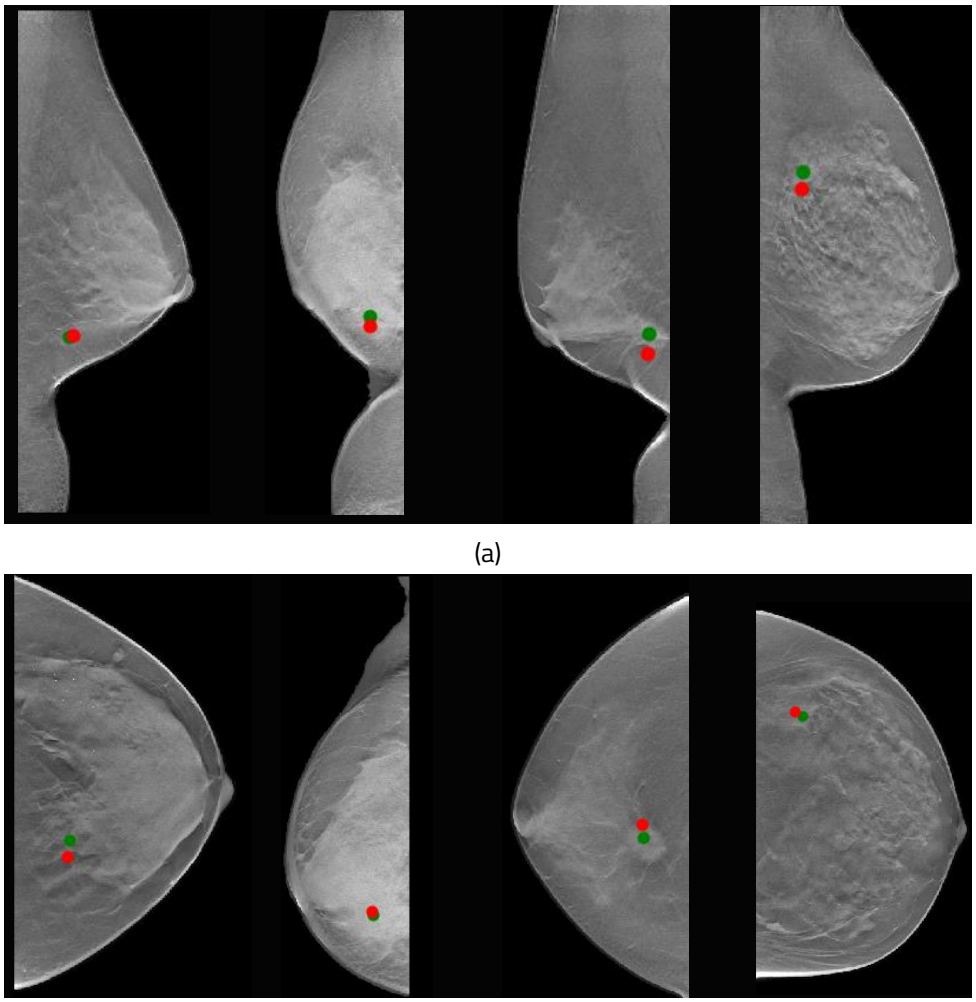
Given the 3D lesion position in one view, the corresponding position in the second view was estimated and compared with the known, annotated, position. For each case, the mapping has been performed and evaluated in both directions: CC to MLO and MLO to CC. Figure 4.11 shows a visual comparison of mass position mapping.

Considering the lesion mapping from both directions, the square root of the average of squared differences between the estimated 3D position of masses and the actual ones was found to be approximately 21.9 mm. The absolute average distance along each axis (Figure 4.12) illustrates that the differences between the actual and matched lesions in the axial plane were marginally longer in the y-direction, as compared to the x-direction.

4.5.2 Two-view Fusion for Mass Detection

For mass detection registration, first, the deep learning-based model proposed in Section 4.2 has been independently applied on the CC and MLO DBT volumes to identify possible masses. Next, the responses generated in one view were used to predict the corresponding locations in the second view using the proposed geometric mass matching criterion.

For testing cases in which both CC and MLO views were available, the performance of the mass detection algorithm was slightly improved by using the joint two-view responses provided by the single-view deep learning-based mass detection model. More specifically, results showed a 13% increase in specificity at a cost of 4% decrease in sensitivity. However, the balance between false-positive rate and true-positive rate can be controlled by altering the threshold at which a finding is considered to be suspicious (high certainty), as outlined in Figure 4.13.



(a)

(b)

Figure 4.11: Results of mass position correlation from (a) CC to MLO and (b) MLO to CC tomosynthesis of four different patients (represented by columns). Mass positions marked with green color represent the landmarks defined by the radiologist and with red are depicted the positions derived with the proposed method.

4.6 Discussions and Conclusion

Herein the automatic mass detection problem in DBT images has been addressed. More specifically, the effectiveness of a fully 2D end-to-end method based on deep convolutional neural network for clinical diagnosis has been investigated. To address the mass detection problem using data-driven solutions, three issues have been addressed: (i) lack of data, (ii) rough annotations, and (iii) false-positive reduction. To mitigate the lack of data implication when dealing with deep networks, a strategy known as fine-tuning has been adopted at two levels. Instead of aiming at training the model from scratch on the small available DBT dataset and to avoid overfitting, the knowledge extracted from a large computer vision dataset has been incorporated into the model as a starting point. The pre-trained model has been first used to learn to segment masses in conventional 2D mammography images. Thereafter, fine-tuned for mass detection on DBT images. To improve and facilitate DBT mass identification, the problem has been formulated in a way that reduces the implication of inaccurate mass boundaries. Moreover, both the training strategy and the loss function have been adapted particularly for the considered dataset. However, because of the nature of the lesion's appearance, a high false-positive rate have been identified in breast imaging related CAD systems. To

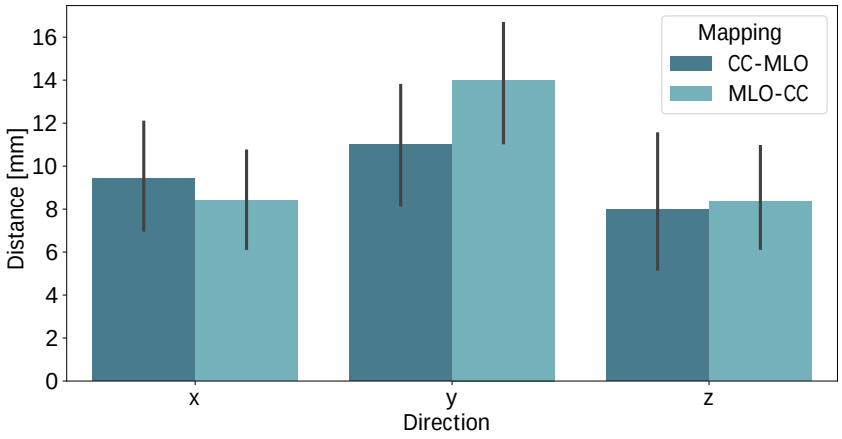


Figure 4.12: Bar graph representation of the absolute average distance between actual and matched 3D masses along the x-axis, y-axis, and z-axis, respectively. Given the presence of landmarks in both CC and MLO DBT views, for each case, the mapping was analyzed from both directions: CC-MLO and MLO-CC. The error bars represent the standard deviation. DBT volumes are characterized by a (0.34 $\times$ 0.34 $\times$ 1) mm resolution.

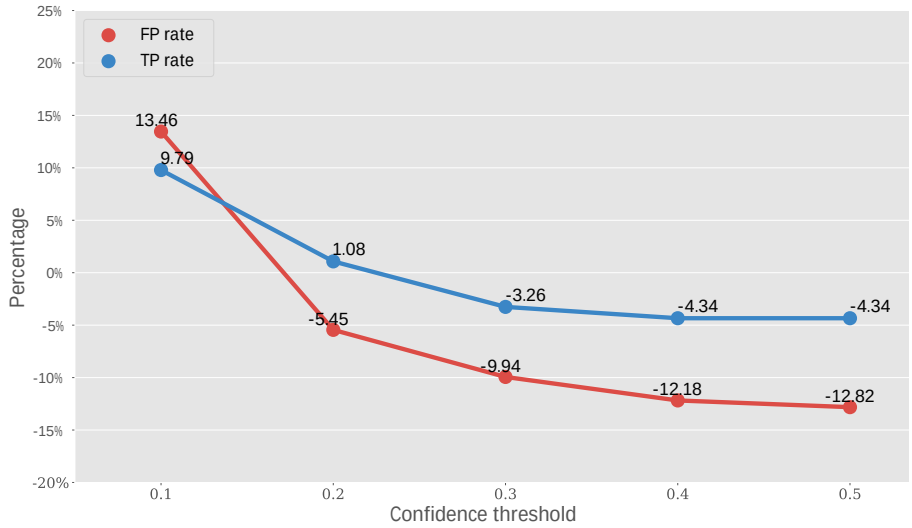


Figure 4.13: Influence of confidence threshold used in location correlation between DBT standard views on TP and FP rate. The threshold controls the value from which a finding is considered suspicious (high certainty). The alteration of these quantities is expressed starting from the performances obtained by the single-view CADe system on mass detection.

reduce the false-positively detected masses by the proposed data-driven solution, a framework that leverages the information provided by the DBT two views of a breast, i.e., CC and MLO, has been further proposed.

In conclusion, the proposed data-driven framework obtained promising results for the challenging problem of mass detection in DBT volumes. Moreover, the two-view assessment framework have shown the potential but also the need of incorporating additional information in the mass detection process. More specifically, it improved the recall rate by eliminating false suspicious findings through view correlations. To further enhance the robustness of deep learning-based mass detection, additional work should be invested in developing an end-to-end fully data-driven framework that combines the two steps.

5. Towards Privacy-Preserving Deep Learning based Medical Applications

5.1 Introduction

Machine learning relies extensively on existing and future patient data to deliver accurate and reliable results [28]. Thus, as access to sensitive plaintext data is required in deep learning-based applications, privacy and security concerns have been raised [29]. Moreover, the currently adopted regulations towards confidentiality guarantees for personal data manipulation (e.g., GDPR in EU, HIPAA in USA) urge for the adoption of more effective privacy-preserving techniques.

Driven by the difficulties that arise in practice when employing deep learning over encrypted data and also by the inefficiency of current solutions, a privacy-preserving solution that increases the efficiency of the encrypted models in real-world applications is investigated. The proposed method enables: (i) computations over rational numbers, (ii) faster operations and, (iii) results comparable to those obtained with the unencrypted model. During the assessment of the proposed solution's feasibility for delivering reliable results, it was shown that the performance does not decrease when deep neural networks operate on data encrypted using the MORE homomorphic encryption scheme. Privacy-preserving deep learning algorithms are applied and evaluated on the classic benchmarking application of digit classification, and on two personalized medicine applications.

5.2 Related Work

5.2.1 Privacy-Preserving Techniques for Machine Learning

A few attempts have been made to address the challenge of data privacy-preserving in machine learning-based analysis through Homomorphic Encryption (HE). This special type of encryption allows data to be encrypted while it is being manipulated. Hence, it aims at keeping the data private by allowing a third party to process the data in the encrypted form without having to reveal the underlying information. By preserving the mathematical structures that underline the data, HE represents a promising solution for guaranteeing privacy while still maintaining full utility.

5.2.2 Homomorphic Encryption

With Gentry's first introduction of a Fully Homomorphic Encryption (FHE) scheme [30], numerous variations of the original strategy were proposed in literature [31]. Most of these schemes are known for their efficiency in terms of security, but they are computationally intensive and only a limited number of operations can be performed before decryption is no longer possible. This clearly restrains their usability in real-world applications. With computations being several orders of magnitude slower than the plaintext counterparts, the accumulated noise that limits the overall number of operations that can be performed and all computations being implemented modulo N, pose a great challenge for the synergy of deep learning and data analysis. While recent advances in HE led to many variants of encryption schemes, no currently available scheme can manipulate rational numbers.

As a consequence, a variant of a matrix-based method, called MORE (Matrix Operation for Randomization or Encryption) [32] was adapted in the current work. As compared to currently studied schemes, in the context of privacy-preserving networks [33], [34], [35], MORE is noise-free (unlimited number of operations can be performed on ciphertext data), and non-deterministic (multiple encryptions of the same message and with the same key result in different ciphertexts). Moreover, both division and multiplication operations can be performed over encrypted data. In order to address the floating-point precision limitation, the MORE encryption scheme was adapted to directly support floating-point arithmetic.

5.3 Matrix-based Data Randomization

Following the MORE approach, a numerical value is encrypted as a matrix and matrix algebra is employed to provide a fully homomorphic behavior, which satisfies both addition and multiplication properties. As a consequence, all operations performed on ciphertext data become matrix operations, e.g., addition of plaintext scalars will result in the addition of ciphertext matrices. The matrix order represents a parameter controlling the trade-off between security and efficiency: by increasing the scheme complexity (i.e., the order of the regular matrix used to encrypt a message) security is improved at a cost of slightly longer running times. MORE encryption scheme can be directly generalized to n by n matrices, however, for simplicity, the 2 by 2 setup is summarized in Table 5.1.

Table 5.1: MORE encryption scheme setup over rational numbers.

Message	Scalar value $m \in \mathbb{R}$
Secret key generation	Invertible matrix $S \in \mathbb{R}^{2\times 2}$
Matrix construction	$M = \left(\begin{smallmatrix} m & 0 \\ 0 & r \end{smallmatrix}\right)$, where $r \in \mathbb{R}$ is a random parameter
Encryption operation	$Encryption(m) = C = SMS^{-1}$
Decryption operation	$Decryption(C) = K = (S^{-1}CS)$
Message recovery	$m = K_{(1,1)}$

5.3.1 Performing Operations over Encrypted Data

The MORE scheme allows for algebraic operations to be performed on encrypted matrices, i.e., given two encrypted matrices $C_1 = SM_1S^{-1}$ and $C_2 = SM_2S^{-1}$, for multiplication

$$C_1C_2 = SM_1S^{-1}SM_2S^{-1} = SM_1M_2S^{-1}, \tag{5.1}$$

which is the encryption of the multiplication M_1M_2, and for addition

$$C_1 + C_2 = SM_1S^{-1} + SM_2S^{-1} = S(M_1 + M_2)S^{-1}. \tag{5.2}$$

Similarly, this property holds true for subtraction and division, but also for operations involving unencrypted scalars, making the scheme fully homomorphic with respect to algebraic operations. A toy example involving computation over ciphertext data is depicted in Figure 5.1.

In real-world applications, including deep learning-based approaches, non-linear (e.g., exponential, logarithmic, square root, etc) functions need to be performed. When an encryption scheme is constrained on using only algebraic operations, the typical approach to support a broader spectrum of operations involves an approximation operation of the non-linear function by finite polynomial series (e.g., truncated Taylor series). The MORE scheme allows for a simple approach for performing such operations. A toy example demonstrating the computation of logarithmic operation over MORE ciphertext data is presented in Figure 5.2. Moreover, Algorithm 5.1 shows how, given any ciphertext

$C \in \mathbb{R}^{2\times 2}$, the two methods can be used to formulate the function $f(x) = \frac{1}{1+e^{-x}}$ defined on $x \in \mathbb{R}$, under the MORE assumptions. This function is known as the logistic sigmoid function and is widely used in neural networks for its non-linear properties.

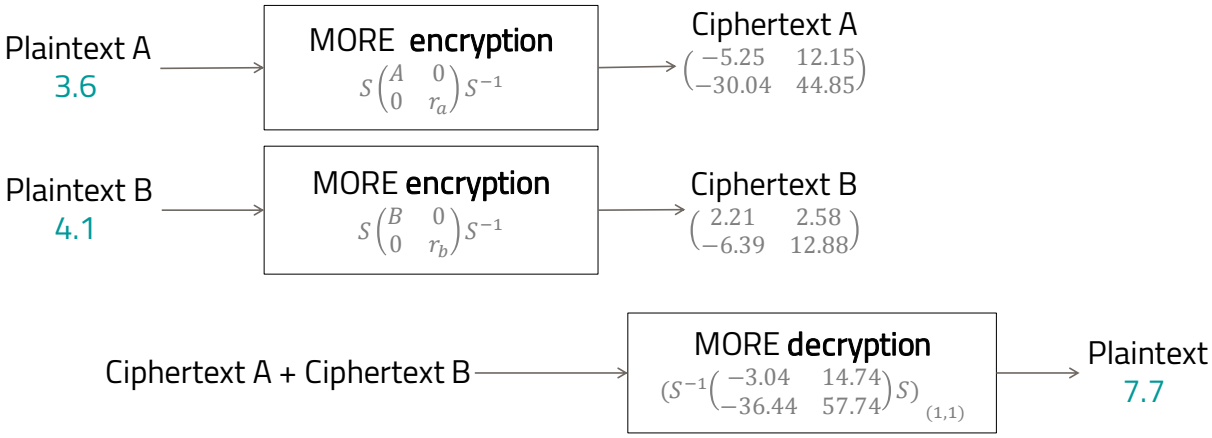


Figure 5.1: Addition example for MORE encryption scheme.

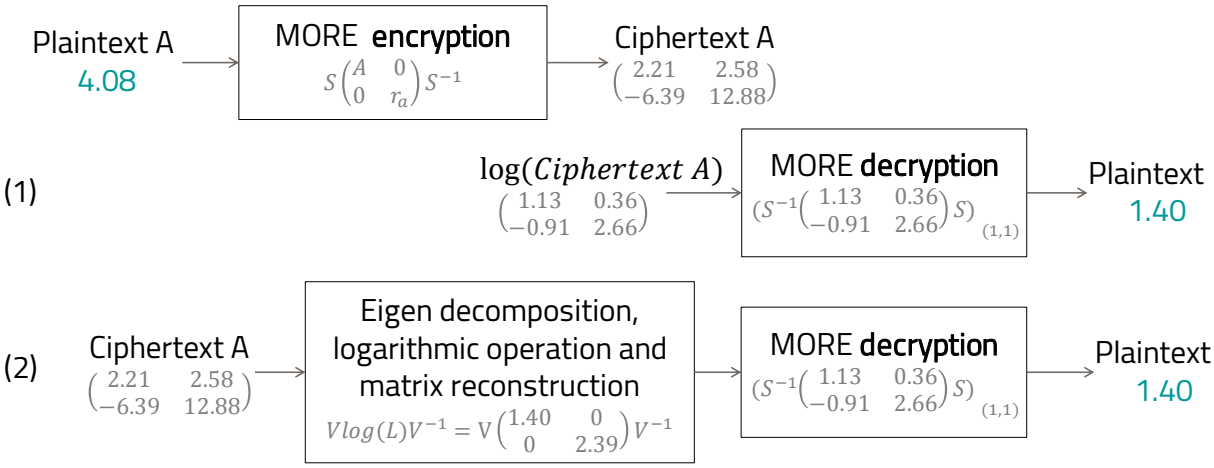


Figure 5.2: Logarithmic example for MORE encryption scheme. Two scenarios are depicted: (1) the straightforward matrix logarithm approach and (2) the eigen decomposition method.

5.4 Deep Neural Networks over Encrypted Data

In this section aspects of privacy-preserving deep neural networks are described. The proposed method is based on the MORE homomorphic encryption scheme and enables both the training and exploitation of classical neural network models directly on homomorphically encrypted data.

5.4.1 Method

The proposed workflow based on HE and deep learning is outlined in Figure 5.3. Before being processed, training data is encrypted with the secret key that is never shared (Algorithm 5.3). Thereafter, the deep learning-based model will have access only to the encrypted version of the data (ciphertext), while the actual data (plaintext) is detached from the processing unit and remains private

Algorithm 5.1 Implementation of the sigmoid function under MORE encryption scheme

Input: Ciphertext $\boldsymbol{C} \in \mathbb{R}^{2\times 2}$
Output: Ciphertext $\boldsymbol{R} \in \mathbb{R}^{2\times 2}$

function Sigmoid($\boldsymbol{C}$) // Using direct matrix operations
    $\boldsymbol{R} \leftarrow \boldsymbol{I}_2 \times (\boldsymbol{I}_2 + \text{MatrixExp}(-\boldsymbol{C}))^{-1}$ // I_2 represents the identity matrix
    return $\boldsymbol{R}$
end function
function Sigmoid($\boldsymbol{C}$) // Using eigen decomposition
    $\boldsymbol{L}, \boldsymbol{V} \leftarrow \text{EigenDecomposition}(-\boldsymbol{C})$
    $\boldsymbol{L}_f \leftarrow \text{Diag}(\text{Exp}(\boldsymbol{L}))$
    $\boldsymbol{C}_{exp} \leftarrow \boldsymbol{V} \times \boldsymbol{L}_f \times \boldsymbol{V}^{-1}$
    $\boldsymbol{R} \leftarrow \boldsymbol{I}_2 \times (\boldsymbol{I}_2 + \boldsymbol{C}_{exp})^{-1}$
    return $\boldsymbol{R}$
end function

on the side of the data provider. Finally, with the homomorphic property underlying the MORE encryption scheme, the direct support for floating-point arithmetic, and with all operations performed inside the network formulated to ensure applicability on ciphertext data, the network can be trained directly on ciphertext data following the classical training pipeline. This results in a model that provides encrypted predictions, which can only be decrypted by the owner of the secret key following Algorithm 5.4. Once the training phase is finalized, the encrypted form of a model can be employed to predict new encrypted instances (inference phase), where input samples are encrypted with the same key as the one used during the training phase. The MORE cryptosystem relies on symmetric keys. Hence, a secret key, generated following the approach presented in Algorithm 5.2, is used for both the encryption of the plaintext data and the decryption of the ciphertext data.

The proposed deep learning-based ciphertext data analysis framework is presented in Algorithm 5.6. Additionally, for the sake of comparison and validation, the pipeline used for plaintext data analysis is provided in Algorithm 5.5. Note that in Algorithm 5.6 all operations performed during training and prediction are formulated in accordance with Section 5.3 and 5.3.1.

Following this approach, privacy is preserved at three levels: (i) during training, when the external party processes directly ciphertexts, (ii) during inference, when the data of the patient remains confidential (the algorithm receives as input ciphertexts and generates outputs as ciphertexts), and (iii)

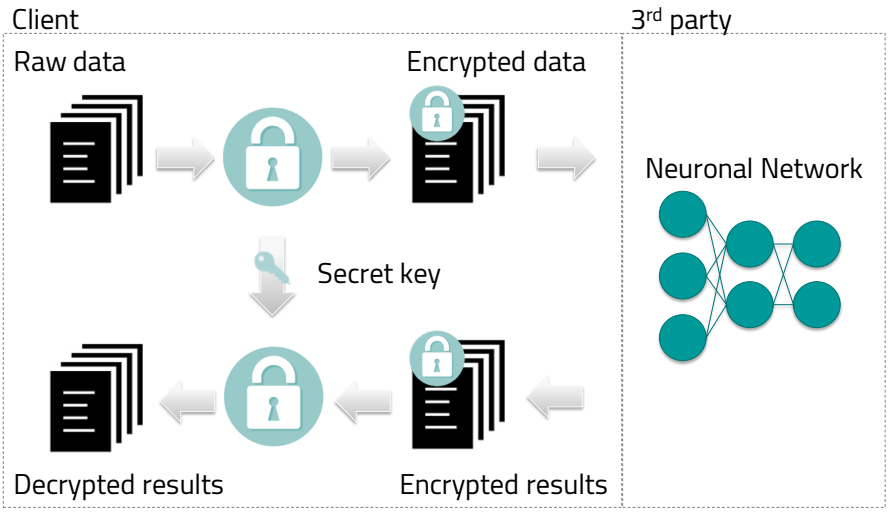


Figure 5.3: Workflow of the proposed privacy-preserving deep learning-based application relying on homomorphic encryption.

Algorithm 5.2 MORE secret key generation

Output: Secret key $\boldsymbol{S} \in \mathbb{R}^{2\times 2}$
function KeyGeneration()
  while True **do**
    $\boldsymbol{S} \leftarrow$ RandomUniform$(size = (2,2), minval, maxval)$
    if det$(\boldsymbol{S}) \neq 0$ **then** // Ensure matrix invertibility
      break
    end if
  end while
  return $\boldsymbol{S}$
end function

Algorithm 5.3 MORE encryption

Input: Plaintext data $m \in \mathbb{R}$
Input: Secret key $\boldsymbol{S} \in \mathbb{R}^{2\times 2}$
Output: Ciphertext $\boldsymbol{C} \in \mathbb{R}^{2\times 2}$
function Encryption($m, \boldsymbol{S}$)
  $\boldsymbol{M} \in \mathbb{R}^{2\times 2} \leftarrow$ zero matrix
  $\boldsymbol{M}(\mathbf{0},\mathbf{0}) \leftarrow m$
  $\boldsymbol{M}(\mathbf{1},\mathbf{1}) \leftarrow$ RandomUniform$(minval, maxval)$
  $\boldsymbol{C} \leftarrow \boldsymbol{S} \times \boldsymbol{M} \times \boldsymbol{S}^{-1}$
  return $\boldsymbol{C}$
end function

Algorithm 5.4 MORE decryption

Input: Ciphertext $\boldsymbol{C} \in \mathbb{R}^{2\times 2}$
Input: Secret key $\boldsymbol{S} \in \mathbb{R}^{2\times 2}$
Output: Plaintext data $m \in \mathbb{R}$
function Decryption($\boldsymbol{C}, \boldsymbol{S}$)
  $\boldsymbol{K} \leftarrow \boldsymbol{S}^{-1} \times \boldsymbol{C} \times \boldsymbol{S}$
  $m \leftarrow \boldsymbol{K}(\mathbf{0},\mathbf{0})$
  return m
end function

Algorithm 5.5 Deep learning-based analysis on plaintext data.

function TrainOnPlaintext()
  $\boldsymbol{X}_{train}, \boldsymbol{Y}_{train} \leftarrow$ LoadDataset()
  $\boldsymbol{X}_{train} \leftarrow$ Normalize$(\boldsymbol{X}_{train})$
  BuildModel()
  Train$(\boldsymbol{X}_{train}, \boldsymbol{Y}_{train})$
  return model
end function
function PredictOnPlaintext()
  $\boldsymbol{X}_{test} \leftarrow$ LoadSamples()
  $\boldsymbol{X}_{test} \leftarrow$ Normalize$(\boldsymbol{X}_{test})$
  LoadModel()
  $\tilde{\boldsymbol{Y}}_{test} \leftarrow$ Predict$(\boldsymbol{X}_{test})$
  return $\tilde{\boldsymbol{Y}}_{test}$
end function

Algorithm 5.6 Deep learning-based analysis on ciphertext data.

function TrainOnCiphertext()
  $\boldsymbol{X}_{train}, \boldsymbol{Y}_{train} \leftarrow$ LoadDataset()
  $\boldsymbol{X}_{train} \leftarrow$ Normalize$(\boldsymbol{X}_{train})$
  $\boldsymbol{S} \leftarrow$ KeyGeneration()
  $\boldsymbol{X}_{train_{enc}} \leftarrow$ Encryption$(\boldsymbol{X}_{train}, \boldsymbol{S})$
  $\boldsymbol{Y}_{train_{enc}} \leftarrow$ Encryption$(\boldsymbol{Y}_{train}, \boldsymbol{S})$
  BuildModel()
  Train$(\boldsymbol{X}_{train_{enc}}, \boldsymbol{Y}_{train_{enc}})$
  return $model_{enc}$
end function
function PredictOnCiphertext()
  $\boldsymbol{X}_{test} \leftarrow$ LoadSamples()
  $\boldsymbol{X}_{test} \leftarrow$ Normalize$(\boldsymbol{X}_{test})$
  $\boldsymbol{S} \leftarrow$ LoadKey()
  $\boldsymbol{X}_{test_{enc}} \leftarrow$ Encryption$(\boldsymbol{X}_{test}, \boldsymbol{S})$
  LoadModel()
  $\tilde{\boldsymbol{Y}}_{test_{enc}} \leftarrow$ Predict$(\boldsymbol{X}_{test_{enc}})$
  $\tilde{\boldsymbol{Y}}_{test} \leftarrow$ Decryption$(\tilde{\boldsymbol{Y}}_{test_{enc}}, \boldsymbol{S})$
  return $\tilde{\boldsymbol{Y}}_{test}$
end function

the external party's deep learning model remains confidential. Consequently, the secure processing of medical data is performed in such a way that the external party cannot derive knowledge from the data, and the user is unable to obtain information regarding the machine learning model.

5.5 Experiments

To validate the proposed method three types of deep learning applications were analyzed: regression, binary and multiclass classification. First, a well-known benchmarking application (digit classification) was addressed, and then the privacy issue was analyzed in two healthcare related applications by training neural network models on encrypted data, (i) to assess whole-body hemodynamics, and (ii) to distinguish coronary artery angiographic views.

The aim of the conducted experiments was not to achieve deep learning-based state-of-the-art results for the proposed problems, but to investigate the possibility of maintaining data privacy while still allowing for computations within a neural network to be successfully performed over the encrypted version of the data.

5.5.1 Problem Formulation

5.5.1.1 MNIST: A Typical Dataset for Neural Networks

A typical problem studied in the context of neural networks is that of classification. More specifically, the problem of image categorization in accordance with the information depicted in the image. The MNIST (Modified National Institute of Standards and Technology) database [36] contains images representing handwritten digits and is typically employed as a reference for benchmarking image classification algorithms (Figure 5.4). The digit recognition problem is framed as predicting the probability of an image belonging to each of the 10 classes (0-9 digits).

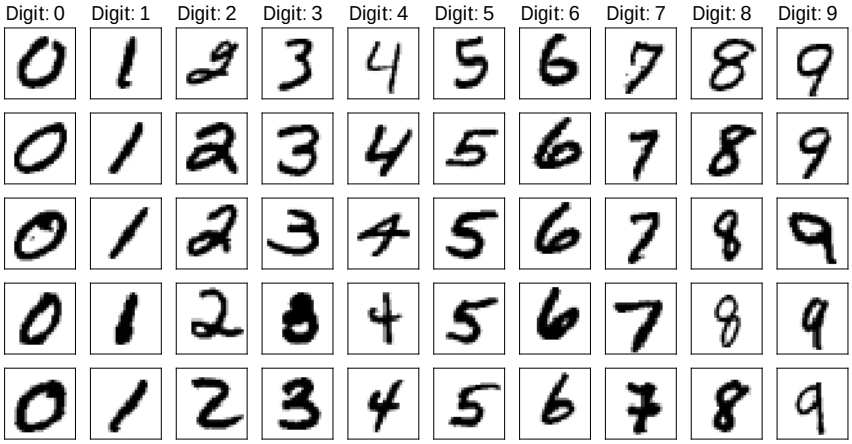


Figure 5.4: Example images from MNIST dataset.

5.5.1.2 Whole-Body Circulation Model

To demonstrate the feasibility of the proposed approach within a personalized medicine application, a hemodynamic model of the cardiovascular system was chosen. More specifically, a whole body circulation (WBC) model. Due to the prohibitive computational cost of spatial blood flow models (three-dimensional models in particular), closed-loop models of the cardiovascular system rely heavily on lumped parameter modeling techniques, which are based on the analogy between hydraulics and electricity. The WBC model employed herein, displayed in Figure 5.5, contains a heart model (left ventricle (LV) and atrium, right ventricle and atrium, valves), the systemic circulation (arteries, capillaries, veins), and the pulmonary circulation (arteries, capillaries, veins) [37].

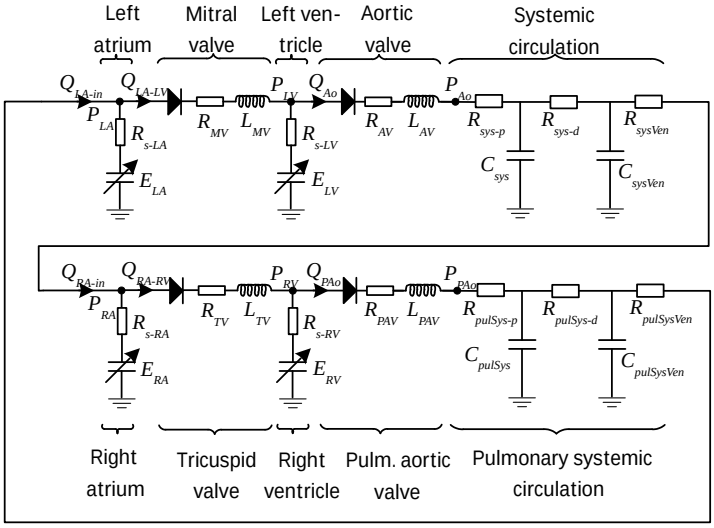


Figure 5.5: Lumped parameter closed-loop model of the cardiovascular system.

The WBC model may be run under patient-specific conditions to compute various clinically relevant measures of interest: arterial resistance, arterial compliance, dead volume of the left / right ventricle, stroke work, ventricular/atrial/arterial elastance, arterial ventricular coupling, pressure-volume loop, etc. However, model parameters need to be personalized to match the patient-specific conditions and state.

The personalization framework employed herein has been previously described in detail [38] and consists of two sequential steps. First, a series of parameters are computed directly, and next, a fully automatic optimization-based calibration method estimates the values of the remaining parameters, ensuring that the personalized computations match the measurements.

The patient-specific input parameters are:

- Systemic circulation: peak aortic systolic pressure, end-diastolic aortic pressure, left ventricular end-systolic and end-diastolic volumes, left ventricular ejection time
- Pulmonary circulation: peak pulmonary artery systolic pressure, end-diastolic pulmonary artery pressure, right ventricular end-systolic and end-diastolic volumes, right ventricular ejection time

The personalized measures of interest determined after running the personalization are:

- Systemic circulation: dead volume of the left ventricle, time at maximum left ventricular elastance, systemic resistance, systemic compliance, ratio of proximal to distal resistance of systemic circulation
- Pulmonary circulation: dead volume of the right ventricle, time at maximum right ventricular elastance, pulmonary resistance, pulmonary compliance, ratio of proximal to distal resistance of pulmonary circulation

While the lumped parameter model is computationally very efficient, its personalization requires hundreds of forward runs, leading to an overall computation time of 30 – 60 seconds for determining the patient-specific measures of interest on a standard desktop hardware configuration. Thus, a model capable of outputting in real-time the measures of interest, that would otherwise be determined using the WBC model, would be a useful tool, even when run under plaintext conditions.

In the context of deep neural network this problem is framed as predicting real-valued quantities from a set of input parameters.

5.5.1.3 X-ray Coronary Angiographies

Invasive X-ray Coronary Angiography (ICA) is a diagnostic imaging procedure that provides important information on the structure and function of the heart, and represents the gold standard in coronary artery disease imaging [39]. ICA enables the assessment of the anatomical severity of coronary stenoses either visually or by computer-assisted Quantitative Coronary Angiography (QCA) [40]. Coronary angiographies are recorded separately and sequentially for the right coronary artery (RCA) and the left coronary artery (LCA) (Figure 5.6).

An important research area in coronary artery disease is the fully automated post-processing of coronary angiographies [41], having as objectives:

- Anatomical assessment: automatically determining the anatomical severity of stenoses.
- Non-invasive functional assessment: automatically computing functional diagnostic indices. [42], [43].
- Reporting: composing medical reports automatically based on the findings in the coronary angiographies.

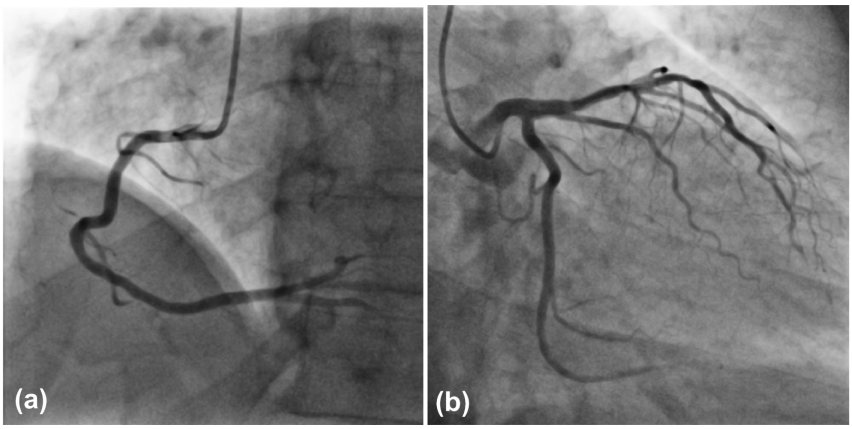


Figure 5.6: (a) Right coronary artery, (b) Left coronary artery.

In many clinical settings based on the use of ICA, automatic LCA / RCA view classification represents an important pre-processing step.

The X-ray coronary angiography view recognition can be formulated as a binary classification problem, where a neural network model learns to predict the probability of an image belonging to the positive class (represented by the value 1).

5.5.2 Ciphertext Database Preparation

To evaluate the performance of the proposed privacy-preserving method, three databases have been used: images of handwritten digits (MNIST), X-ray coronary angiographies, and synthetically generated WBC samples. A brief overview of these databases is given in Table 5.2.

To address the challenge of privacy-preserving computations and to evaluate the use of deep neural network models over encrypted data, for each dataset the input samples, i.e., image or feature vectors, were encrypted following the MORE encryption strategy, as described in Algorithm 5.3. A ciphertext representation of a digit extracted from the MNIST database is depicted in Figure 5.7. Similarly, the target values, i.e., class labels or real-valued quantities, were also encrypted.

Table 5.2: Overview of databases used for experimental evaluation.

	MNIST	WBC	Angio
Number of training samples	50000	7000	1996 (7984*)
Number of validation samples	10000	1000	680
Number of testing samples	10000	2000	702

*after augumentation

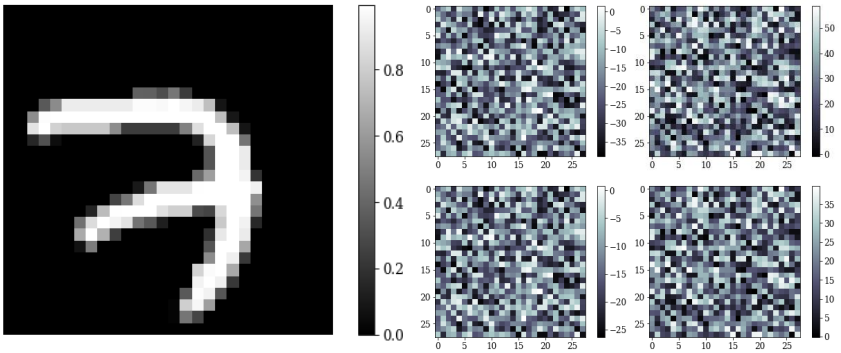


Figure 5.7: Plaintext digit image with pixels in interval $p \in [0, 1]$ shown alongside ciphertext components.

5.5.3 Deep Neural Network Models Architecture

To assess the feasibility and effectiveness of deep neural networks to operate directly on homomorphically encrypted data three experiments were conducted by training: (i) a CNN for digit recognition on encrypted handwritten images, (ii) a traditional FCNN for real-time hemodynamic analysis, where both the input feature vector and the ground truth outputs were encrypted, and (iii) a CNN for encrypted X-ray coronary angiographies view classification. For a comparison of model performance and convergence, training was also performed on the counterpart models on plaintext data.

Although more efficient alternative deep neural network models (e.g., improved activation functions, greater depth) can be adopted to ensure better convergence and superior performance, herein the purpose of the experiments was to asses the correctness and effectiveness of different deep neural network models operating on ciphertext data, as compared to the counterpart models trained on plaintext data.

5.5.3.1 Deep Neural Network for Handwritten Digit Classification

Starting from the latest results obtained by CNN models on the MNIST digit recognition task, a CNN was employed on encrypted input-output value pairs. The topology of the proposed privacy-preserving CNN is described in Table 5.3.

5.5.3.2 Deep Neural Network for Real-time Hemodynamic Analysis

Given the nature of the input data, i.e., information represented as a feature vector, and driven by the need to model the decision of the network based on a global dependency between input features, a fully connected neural network with 3 hidden layers was employed. The topology of the proposed privacy-preserving FCNN is described in Table 5.4.

Table 5.3: CNN-MNIST: The topology of the CNN for handwritten digits classification.

Layers	Parameters	Dimensions
Input	-	**(1,28,28)**
Convolution	(8,3,3)	(8,28,28)
Activation (Sigmoid)	-	-
Average Pooling	(2,2)	(8,14,14)
Convolution	(16,3,3)	(16,14,14)
Activation (Sigmoid)	-	-
Average Pooling	(2,2)	(16,7,7)
Flatten	-	(784,)
Fully Connected	100	(100,)
Activation (Sigmoid)	-	-
Fully Connected	10	(10,)
Activation (Softmax)	-	-

Table 5.4: FCNN-WBC: The topology of the FCNN for hemodynamic analysis.

Layers	Parameters	Dimensions
Input	-	**(9,)**
Fully connected	40	(40,)
Activation (Tanh)	-	-
Fully connected	40	(40,)
Activation (Tanh)	-	-
Fully connected	40	(40,)
Activation (Sigmoid)	-	-
Fully connected	12	(12,)
Activation (Linear)	-	-

5.5.3.3 Deep Neural Network for View Classification in X-ray Coronary Angiography

Motivated by the latest results in data-driven image-based analysis, a deep CNN was adopted to solve the coronary angiography image recognition task. The topology of the proposed privacy-preserving CNN is described in Table 5.5.

5.6 Results

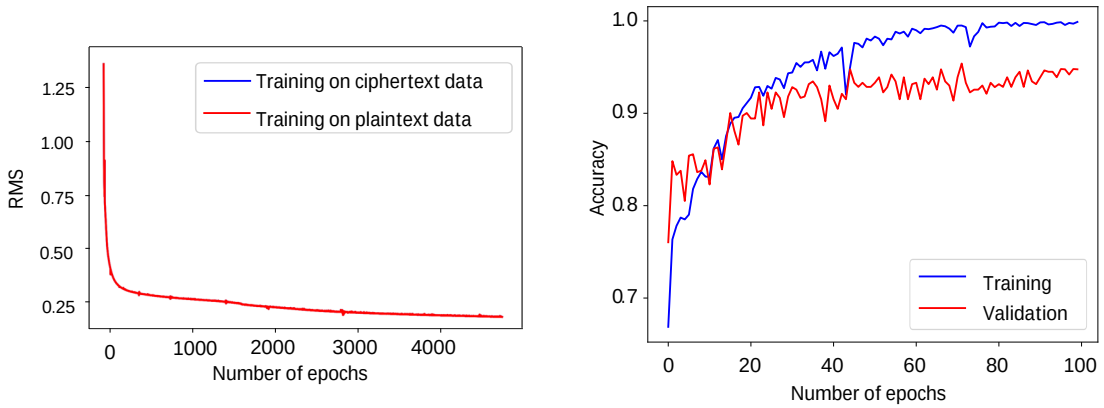
5.6.1 Performance

To showcase the ability of the network to learn from ciphertext data, the training loss for the regression task, as resulted after decryption, is depicted in Figure 5.8a. Similarly, the evolution of the training and validation accuracy of the privacy-preserving CNN model fed with encrypted X-ray coronary angiographies, obtained after decryption, is depicted in Figure 5.8b.

The training evolution demonstrates the capability of the proposed method to preserve the correctness of the computations. Moreover, after decryption, the parameters learned by the model when trained on ciphertext data were found to be identical up to machine precision to those learned by the unencrypted model.

Table 5.5: CNN-Angio: The topology of the CNN for view classification in X-ray coronary angiographies.

Layers	Parameters	Dimensions
Input	-	**(1,256,256)**
Convolution	(4,3,3)	(4,256,256)
Activation (Sigmoid)	-	-
Average Pooling	(2,2)	(4,128,128)
Convolution	(8,3,3)	(8,128,128)
Activation (Tanh)	-	-
Average Pooling	(2,2)	(8,64,64)
Convolution	(16,3,3)	(16,64,64)
Activation (Tanh)	-	-
Average Pooling	(2,2)	(16,32,32)
Convolution	(32,3,3)	(32,32,32)
Activation (Tanh)	-	-
Average Pooling	(2,2)	(32,16,16)
Flatten	-	(8192,)
Fully connected	64	(64,)
Activation (Tanh)	-	-
Dropout	25%	-
Fully connected	-	(1,)
Activation (Sigmoid)	-	-



(a) WBC parameter estimation networks

(b) Angiographic view classification network

Figure 5.8: (a) Evolution of the training loss for encrypted and unencrypted networks: differences between the learning curves, caused by floating-point arithmetic, are unnoticeable. (b) Evolution of the accuracy when training on ciphertext data.

5.6.1.1 MNIST Binary Classification

The default metric used to assess the performance of a classifier on the MNIST dataset is given by the absolute accuracy of the classification models, i.e., the percentage of correctly labeled digit images. The unencrypted network achieved a classification accuracy of 98.2% on the testing dataset,

which was preserved by the encrypted network.

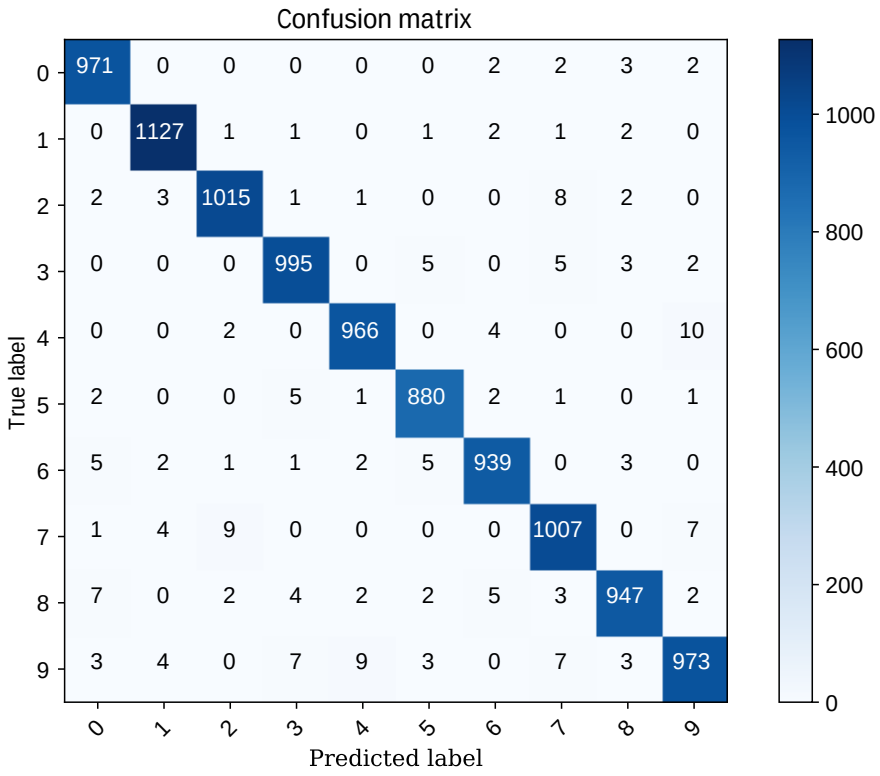


Figure 5.9: Confusion matrix of the MNIST digit classification task on the test set. The number on the diagonal indicates the number of correctly classified images, while the rest represent the misclassified ones.

Table 5.6: Precision, recall, and F1-score, of the deep neural network-based MNIST digit classification.

Digit	Precision (%)	Recall (%)	F1-score (%)
0	97.9	99.0	98.5
1	98.8	99.2	99.0
2	98.5	98.3	98.4
3	98.1	98.5	98.3
4	98.4	98.3	98.4
5	98.2	98.6	98.4
6	98.4	98.0	98.2
7	97.3	97.9	97.6
8	98.3	97.2	97.7
9	97.5	96.4	97.0
Average	**98.1**	**98.1**	**98.1**

Being a multi-class classification problem, the evaluation metrics were computed following the one-vs-rest strategy. More specifically, to compute the metrics each label was individually considered positive while the others were set as being negative. The precision, recall, and f1-score for each

digit class are reported in Table 5.6. To evaluate the digit recognition performance of the proposed CNN model the confusion matrix was computed and is displayed in Figure 5.9.

5.6.1.2 Hemodynamic Analysis

To evaluate the capability of the deep neural network model to estimate the outputs of the whole-body circulation model, the mean absolute relative error and the Pearson correlation were computed and the results are displayed in Table 5.7. Scatter plots of the measured versus predicted parameters, having the highest and lowest correlation coefficient, are presented in Figure 5.10. The first scatter plot displays the results of the neural network model obtained for estimating the ratio of proximal to distal resistance in the systemic circulation. The latter presents the results for systemic resistance prediction.

Table 5.7: Results of the deep neural network for real-time hemodynamic analysis on the testing dataset.

Circulation	Parameters	MAPE (%)	Pearson correlation (%)
Systemic	Dead volume	7.03	0.9997
	Time at max. elastance	0.13	0.9995
	Resistance	0.17	0.9999
	Compliance	2.45	0.9867
Pulmonary	Dead volume	9.88	0.9991
	Time at max. elastance	0.10	0.9994
	Resistance	0.32	0.9998
	Compliance	0.67	0.9983

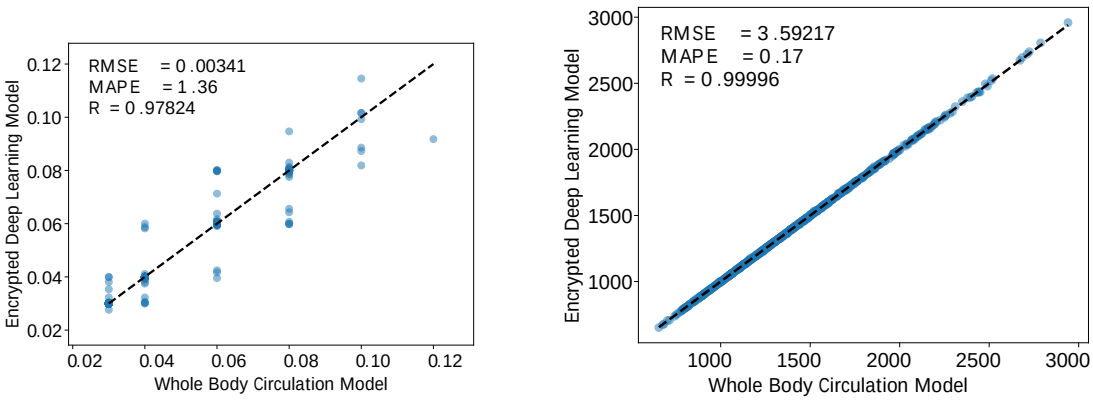


Figure 5.10: Predicted versus ground truth (a) ratio of proximal to distal resistance in the systemic circulation, and (b) systemic resistance.

5.6.1.3 X-ray Coronary Angiographies Classification

To assess the accuracy of the coronary angiography view recognition model, the obtained ROC (receiver operator characteristic) curve is displayed in Figure 5.12. Table 5.8 lists the precision, recall, and f1-score for both the LCA and RCA labels. Figure 5.11 displays the confusion matrix, portraying measures of association between the true labels and the deep neural network predictions of LCA and RCA.

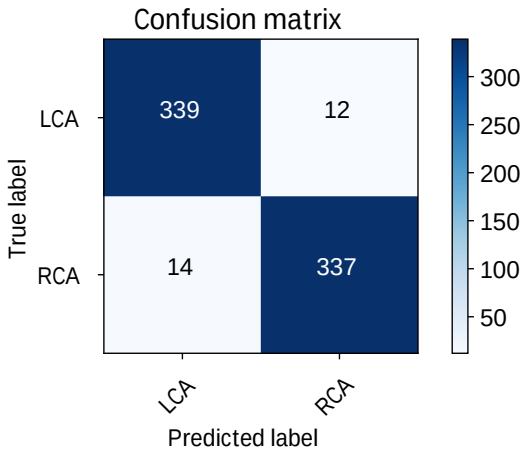


Figure 5.11: Confusion matrix of the X-ray coronary angiography view classifier.

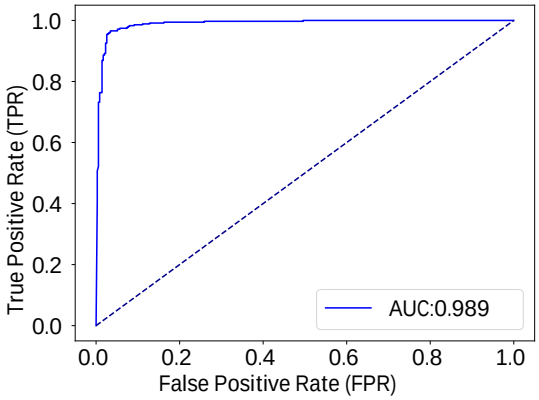


Figure 5.12: ROC curve of the view classification task in X-ray coronary angiography.

Table 5.8: Precision, recall, and F1-score, of deep neural network for hemodynamic analysis.

Label	Precision (%)	Recall (%)	F1-score (%)
LCA	96.0	96.5	96.3
RCA	96.5	96.0	96.2
Average	**96.2**	**96.2**	**96.2**

In the angiographic view classification use case, the CNN network trained on ciphertext data classified 96.2% of the samples correctly when evaluated on the held-out testing angiographies. When compared to the unencrypted model, accuracy was identical.

5.6.2 Execution Time

A detailed comparison of the runtime for each of the medical applications is given in Table 5.9. Although deep learning models run directly on MORE homomorphically encrypted data are significantly slower (up to one order of magnitude) during both training and inference, the scheme is currently outstandingly faster compared to classic FHE schemes where the difference is of around 6 to 7 orders of magnitude, even when performing very basic algebraic operations.

5.7 Discussion and Conclusions

In the past few years, the raised concern for protecting the privacy of sensitive medical data, while still encouraging the delivery of personalized medicine solutions, increased the focus on enabling privacy-preserving computations inside deep neural networks.

The proposed solution aims at ensuring privacy by incorporating a data encryption mechanism and delivering reliable results, to be used in clinical workflows. The applicability of incorporating the

Table 5.9: Runtime analysis: mean values and standard deviation of the encrypted and plaintext networks for the two personalized medicine use cases.

Task	Operation	Runtime (s) on cipertext data	Runtime (s) on plaintext data	Encrypted - Unencrypted ratio
Angiographic view classification	Training (1 epoch)	1075.47±45.54	34.48±1.12	31.19
	Inference (702 images)	26.36±1.98	0.8±0.06	32.95
WBC hemodynamic analysis	Training (1 epoch)	0.66±0.09	0.021±0.001	31.4
	Inference (2000 samples)	0.102±0.01	0.006±0.0009	17

MORE encryption scheme into deep learning models has been showcased by tackling three different problems: digit recognition, whole body hemodynamic analysis, and coronary angiography view classification. Both the training and the inference phase were addressed, and it was shown that both can be performed on encrypted data. Two main quantities were tracked: (i) the difference between results obtained using the classical approach (with no encryption) and results obtained using encryption, and (ii) computation time differences between the two scenarios. It was demonstrated that the accuracy of the encrypted model is statistically not discernible from that of the unencrypted model, and that, by following the proposed strategy, computations over ciphertext data are only slightly more costly than the ones performed on plaintext data.

In conclusion, experiments showed that employing the MORE fully homomorphic encryption scheme as a privacy-preserving mechanism enabled the application of deep learning models on encrypted data without compromising the accuracy at all. Although the runtime increased by more than one order of magnitude, the encrypted models are still outputting results in a reasonable amount of time. With its direct support for computations over rational numbers, and the ability to perform operations without adding noise, the scheme becomes eligible for more complex deep learning models. Note that the scheme allows for a trade-off between security and efficiency: by increasing the scheme complexity (i.e., the order of the regular matrix used to encrypt a message) security is improved at a cost of slightly longer runtimes.

6. Final Conclusions

6.1 Conclusions

When use by physicians, data-driven solutions, and specifically deep neural networks, can bring major advances in the delivery of care by improving health (e.g., earlier or improved diagnostics, effective treatment plans, improved disease prevention, personalized medicine, etc.), patient experience (e.g., lower exposure time, faster diagnosis, non-invasive procedures, etc.), and reducing the cost of care. Due to the shift towards electronic health records and the current routine clinical workflows, a patient's health record information may include hundreds of interrelated and interdependent clinical, genomic, and imaging data, collected through the years. To alleviate the difficulty of analyzing, interpreting, and synthesizing complex medical data, large-scale data-driven solutions have been proposed as means of enabling broader insights into the data.

The aim of the research described herein has been to investigate the potential of deep learning-based models to positively impact healthcare. Specifically, to develop, implement, test, and verify data-driven solutions for medical imaging analysis. Although machine learning is predominantly integrated into the development of computer-aided detection and diagnosis systems herein, the ultimate goal is to infuse data-driven solutions in the process of medical imaging analysis starting from image acquisition, continuing with secure patient health data sharing and up to the final step of image interpretation. Hence, the thesis lays out the basis towards the replacement of classical approaches with fully automated machine learning-based solutions for reduced diagnostic time and improved accuracy.

Considering first the deep learning-based system for inverse problems developed, implemented, and tested in Chapter 3: the reconstruction solution has been shown to produce promising results, as validated by experienced radiologists, for full-dose tomography reconstruction, starting from low-dose measurements. Moreover, the computed tomography image reconstruction is obtained in a runtime suitable for a routine clinical setting. Similarly, a fully automatic breast mass detection system is developed and validated in Chapter 4. The deep convolutional neural network-based solution seeks to reduce the workload and improve physician efficiency in interpreting 3D digital breast images, which is known to pose great challenges in clinical routines.

Although promising results have been achieved over the years in imaging, it is still a long road ahead for such systems to become completely reliable. A clear bottleneck is given by the current strict regulations towards data protection. Despite the fact that an abundant amount of data is collected in day-to-day clinical care, it is locked inside hospitals firewalls. Hence, one of the greatest challenges in the biomedical industry is to develop and provide personalized medical care solutions without disclosing sensitive patient health-related information. However, these two requirements are mutually exclusive. To provide reliable personalized medical care solutions, researchers have to rely extensively on existing patient data, which is challenging to make available. Additionally, there is a clear need for a joint effort to aggregate data acquired at different healthcare facilities, to continuously adapt and improve learning-based solutions. However, this implies that sensitive personal data is shared. Chapter 5 provides a possible solution for guaranteeing data confidentiality for deep learning-based personal data manipulation. The reported results indicate that the proposed solution has great potential: (i) computational results are indistinguishable from those obtained with the unencrypted variants of the deep learning-based applications, and (ii) runtimes increase only

marginally. It has been introduced as a possible framework for facilitating joint collaboration between researchers towards faster development of supporting systems for medical decision making.

6.2 Personal Contributions

The personal thesis contributions fall into three major categories:

- Deep learning-based medical imaging reconstruction;
- Deep learning-based diagnosis;
- Privacy-preserving deep learning.

6.2.1 Deep Learning-based Medical Imaging Reconstruction

One of the most active research areas in computed tomography (CT) is to devise strategies to reduce radiation exposure, while maintaining high image quality, required for accurate diagnosis. The recent advancements offered by deep learning-based data-driven approaches for solving inverse problems in biomedical imaging have led to the development of an alternative method for producing high-quality reconstructed images from low-dose CT data. While most of the reconstruction approaches tackle the problem from a post-processing perspective, herein, inspired by the idea of unfolding a proximal gradient descent optimization algorithm to finite iterations, and replacing the proximal terms with trainable deep artificial neural networks, an end-to-end solution for reconstructing full-dose tomographic images directly from low-dose measurements is proposed.

To summarize, the main contributions in this part are:

- design and development of a novel end-to-end framework for inverse-problems in medical imaging integrating deep convolutional neural networks, the physical model of CT image formation, and adversarial training;
- design and development of an improved learning-based iterative reconstruction framework incorporating the generative adversarial network with Wasserstein distance and a contextual loss for higher-quality reconstructed images that account for human perception;
- design, validation and application of the framework on a clinically-realistic scenario;
- design and development of a clinically realistic evaluation tool for physicians-based assessment of the CT image reconstruction quality;

6.2.2 Deep Learning-based Diagnosis

Motivated by recent advances in deep learning across a wide range of areas in healthcare and by the need for advanced computer-aided diagnosis systems, a breast imaging analysis framework based on data-driven models (proven to be superior to classical machine learning techniques) is introduced. More specifically, a deep convolutional neural network model is trained in a supervised manner to highlight suspicious regions in digital breast tomosynthesis (DBT) images. To cope with the high variance in lesion appearance and the uncertain boundaries, the mass detection problem is cast as a confidence map detection problem (e.g. heat map centred on the lesion location), instead of defining the location though the bound box coordinates. Moreover, a training strategy is adopted to cope with the imbalanced class distribution that appears in the data, and to facilitate small mass detection. To alleviate the difficulties that arise when dealing with small-sized datasets, an existing publicly available mammography dataset is adopted to pre-train the model, and improve the generalization capability of the lesion detection solution. Additionally, a novel mass matching framework is proposed to improve detection, and reduce the false-positive findings.

To summarize, the main contributions of this part are:

- design and development of an end-to-end framework for mass detection in digital breast tomosynthesis images;
- exploration of the confidence map detection formulation for the mass identification problem;
- design and development of a training strategy to exploit the 3D nature of the data;
- design and development of a fine-tuning strategy to improve model generalization;
- design and development of a novel post-processing framework based on mass registration across breast views for improved detection;
- design, application, and validation of the framework using an in-house imaging database;
- experimental study on assessing the performance.

6.2.3 Privacy-Preserving Deep Learning

Despite the potential of machine learning in enabling personalized medical care applications, the adoption of deep learning-based solutions in clinical workflows has been hindered in many cases by the strict regulations concerning the privacy of patient health data. A solution that relies on Fully Homomorphic Encryption (FHE), particularly on the MORE (Matrix Operation for Randomization or Encryption) scheme, is proposed as a mechanism for enabling computations on sensitive health data, without revealing the underlying data. The chosen variant of the encryption scheme allows for the computations in the neural network model to be performed directly on floating-point numbers, while incurring a reasonably small computational overhead. For feasibility evaluation, the MNIST digit recognition task is used to demonstrate that deep learning can be performed on encrypted data without compromising the accuracy. To evaluate the suitability of the proposed method for healthcare applications, a model is first trained on encrypted data to estimate the outputs of a whole-body circulation (WBC) hemodynamic model, and then a solution for classifying encrypted X-ray coronary angiography medical images is provided. The experimental results underline the potential of the proposed approach to outperform current solutions, by delivering results comparable to those obtained with the unencrypted deep learning-based solutions, in a reasonable amount of time. The security aspects of the encryption scheme are analyzed, and it is shown that, even though the chosen encryption scheme favors performance and utility at the cost of weaker security, it can still be used in certain practical applications, while still, significant limitations remain to be solved in future work. Lastly, possible solutions to mitigate the proposed scheme vulnerability are studied and a follow-up scheme with an additional obfuscation layer, Hybrid More, is proposed. Results show that the scheme security can be strengthened but further improvements are needed to limit the growth of the introduced noise.

To summarize, the main contributions of this part are:

- design, development and evaluation of a secure noise-free homomorphic encryption scheme that allows for operations to be performed directly on floating-point numbers;
- design and development of a generic privacy-preserving deep learning library operating on homomorphically encrypted data;
- validation and application of the privacy-preserving deep learning framework on a benchmark digit recognition task;
- validation and application of the privacy-preserving deep learning framework on whole-body hemodynamic analysis;
- validation and application of the privacy-preserving deep learning framework on coronary angiography image analysis;

- experimental study for assessing the performance and security;
- design, development and evaluation of a novel homomorphic encryption scheme (Hybrid More) that combines MORE encryption with an additional obfuscation layer for improved security.

6.3 Dissemination of Research Results

The work undertaken during the PhD studies led to a number of publications in international scientific journals and conference proceedings. Specifically, 5 research papers have been publish as first author and key contributions have been made to another 4, as follows:

- **Vizitiu, A.**, Nita, C., Puiu, A., Suciu, C., Itu, L., Applying Deep Neural Networks over Homomorphic Encrypted Medical Data, Computational and Mathematical Methods in Medicine, 2020;
- **Vizitiu, A.**, Nita, C., Puiu, A., Suciu, C., Itu, L., Towards Privacy-Preserving Deep Learning based Medical Imaging Applications, IEEE International Symposium on Medical Measurements and Applications (MeMeA), 2019;
- **Vizitiu, A.**, Nita, C., Puiu, A., Suciu, C., Itu, L., Privacy-Preserving Artificial Intelligence: Application to Precision Medicine, 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019;
- **Vizitiu, A.**, Puiu, A., Reaungamornrat, S., Itu, L., Data-Driven Adversarial Learning for Sinogram-Based Iterative Low-Dose CT Image Reconstruction, 23rd International Conference on System Theory, Control and Computing (ICSTCC), 2019;
- Danu, M., Nita, C., **Vizitiu, A.**, Suciu, C., Itu, L., Deep Learning-based Generation of Synthetic Blood Vessel Surfaces, 23rd International Conference on System Theory, Control and Computing (ICSTCC), 2019;
- **Vizitiu, A.**, Nita, C., Itu, L., Homomorphic Encryption in Deep Learning-based Applications for Healthcare Data Analysis, Transylvanian Machine Learning Summer School (TMLSS), 2018;
- Ciusdel, C., **Vizitiu, A.**, Moldoveanu, F., Suciu, C., Itu, L., Towards Real Time Machine Learning-based Estimation of Fracture Risk in Osteoporosis Patients, International Conference on Optimization of Electrical and Electronic Equipment (OPTIM), 2017 and International Aegean Conference on Electrical Machines and Power Electronics (ACEMP), 2017;
- Suciu, C., Itu, L., Nita, C., **Vizitiu, A.**, Stroia, I., Lazăr, L., Gîrbea, A., Foerster, U., Mihalef, V., Patient-specific Hemodynamic Computations: Application to Personalized Diagnosis of Cardiovascular Pathologies, pp. 177-227, 2017;
- Ciusdel, C., **Vizitiu, A.**, Moldoveanu, F., Suciu, C., Itu, L., Towards Deep Learning-based Estimation of Fracture Risk in Osteoporosis Patients, 40th International Conference on Telecommunications and Signal Processing (TSP), 2017.

Moreover, the research results have been promoted at a series of conferences and venues and attracted lots of attention and discussions. Consequently, 3 awards have been earned:

- **"Best Poster Award"** for the work conducted towards privacy-preserving deep learning at a Machine Learning summer school conference organized by Google DeepMind;
- **"Innovation Radar Prize 2019"** at "Industrial & Enabling Tech" category where, as part of the EU-funded research project H2020 "My Health – My Data", the high-potential innovation of the privacy-preserving framework has been recognized;

- **"Best Paper Award for Ph.D. students"** for the CT deep learning-based reconstruction paper presented at IEEE ICSTCC conference.

These awards recognize the value of the thesis, but also the research performance and the potential impact of this work on the future of the medical industry.

References

[1] Arthur L. Samuel. "Some Studies in Machine Learning Using the Game of Checkers". In: IBM Journal of Research and Development 3 (1959), pp. 210–229.

[2] Frost and Sullivan. From $600M to $6 billion, artificial intelligence systems poised for dramatic market expansion in healthcare. 2016. url: https://ww2.frost.com/news/press-releases/%20600-m-6-billion-artificial-intelligence-systems-poised-dramaticmarket-expansion-healthcare.

[3] Anamaria Vizitiu, Andrei Puiu, Sureerat Reaungamornrat, and Lucian M Itu. "Data-Driven Adversarial Learning for Sinogram-Based Iterative Low-Dose CT Image Reconstruction". In: 2019 23rd International Conference on System Theory, Control and Computing (ICSTCC) (2019), pp. 668–674.

[4] Anamaria Vizitiu, Cosmin Ioan Nita, Andrei Puiu, Constantin Suciu, and Lucian Mihai Itu. "Privacy Preserving Artificial Intelligence: Application to Precision Medicine". In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2019), pp. 6498–6504.

[5] Anamaria Vizitiu, Cosmin Ioan Nita, Andrei Puiu, Constantin Suciu, and Lucian Mihai Itu. "Towards Privacy Preserving Deep Learning based Medical Imaging Applications". In: 2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA) (2019), pp. 1–6.

[6] Anamaria Vizitiu, Cosmin Ioan Nita, Andrei Puiu, Constantin Suciu, and Lucian Mihai Itu. "Applying Deep Neural Networks over Homomorphic Encrypted Medical Data". In: Computational and Mathematical Methods in Medicine (2020).

[7] Warren Sturgis McCulloch and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity". In: Bulletin of Mathematical Biology 52 (1988), pp. 99–115.

[8] Hao Yan, Laura Cervino, Xun Jia, and Steve B. Jiang. "A comprehensive study on the relationship between the image quality and imaging dose in low-dose cone beam CT." In: Physics in medicine and biology 57 7 (2012), pp. 2063–80.

[9] Ge Wang, Jong Chul Ye, Klaus Mueller, and Jeffrey A. Fessler. "Image Reconstruction is a New Frontier of Machine Learning". In: IEEE Transactions on Medical Imaging 37 (2018), pp. 1289–1296.

[10] Jonas Adler and Ozan Oktem. "Learned Primal-Dual Reconstruction". In: IEEE Transactions on Medical Imaging 37 (2018), pp. 1322–1332.

[11] Michael T. McCann, Kyong Hwan Jin, and Michael Unser. "A Review of Convolutional Neural Networks for Inverse Problems in Imaging". In: ArXiv abs/1710.04011 (2017).

[12] Jonas Adler and Ozan Öktem. "Solving ill-posed inverse problems using iterative deep neural networks". In: CoRR abs/1704.04058 (2017).

[13] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. "Loss Functions for Image Restoration With Neural Networks". In: IEEE Transactions on Computational Imaging 3 (2017), pp. 47–57.

[14] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. "Perceptual Losses for Real-Time Style Transfer and Super-Resolution". In: ECCV. 2016.

[15] Ishaan Gulrajani, Faruk Ahmed, Martín Arjovsky, Vincent Dumoulin, and Aaron C. Courville. "Improved Training of Wasserstein GANs". In: Conference on Neural Information Processing Systems (NIPS). 2017.

[16] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. "Image-to-Image Translation with Conditional Adversarial Networks". In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017), pp. 5967–5976.

[17] Karen Simonyan and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: International Conference on Learning Representations (ICLR) (2015).

[18] Cynthia H. McCollough et al. "Low-dose CT for the detection and classification of metastatic liver lesions: Results of the 2016 Low Dose CT Grand Challenge." In: Medical physics 44 10 (2017), e339–e352.

[19] Independent UK Panel on Breast Cancer Screening. "The benefits and harms of breast cancer screening: an independent review". In: The Lancet 380 (2012), pp. 1778–1786.

[20] Robyn Gartner Roth, Andrew D. A. Maidment, Susan P. Weinstein, Susan Orel Roth, and Emily F. Conant. "Digital breast tomosynthesis: lessons learned from early clinical implementation." In: Radiographics : a review publication of the Radiological Society of North America, Inc 34 4 (2014), E89–102.

[21] Junqiang Lei, Pin Fan Yang, Li Zhang, Yinzhong Wang, and Kehu Yang. "Diagnostic accuracy of digital breast tomosynthesis versus digital mammography for benign and malignant lesions in breasts: a meta-analysis". In: European Radiology 24 (2013), pp. 595–602.

[22] Jun Wei, Heang-Ping Chan, Chuan Zhou, Y. Wu, Berkman Sahiner, Lubomir M. Hadjiiski, Marilyn A. Roubidoux, and Mark A. Helvie. "Computer-aided detection of breast masses: four-view strategy for screening mammography." In: Medical physics 38 4 (2011), pp. 1867–76.

[23] Md Atiqur Rahman and Yuhuai Wang. "Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation". In: International Symposium on Visual Computing (ISVC). 2016.

[24] Jeroen Bertels, Tom Eelbode, Maxim Berman, Dirk Vandermeulen, Frederik Maes, Raf Bisschops, and Matthew B. Blaschko. "Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory and Practice". In: ArXiv abs/1911.01685 (2019).

[25] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: Conference on Neural Information Processing Systems (NIPS). 2012.

[26] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation". In: 2016 Fourth International Conference on 3D Vision (3DV) (2016), pp. 565–571.

[27] Edward A. Sickles, Willi Weber, Hazel Galvin, Steven H. Ominsky, and Richard A. Sollitto. "Baseline screening mammography: one vs two views per breast." In: AJR. American journal of roentgenology 147 6 (1986), pp. 1149–53.

[28] Ziad Obermeyer and Ezekiel J. Emanuel. "Predicting the Future - Big Data, Machine Learning, and Clinical Medicine." In: The New England journal of medicine 375 13 (2016), pp. 1216–9.

[29] Reza Shokri and Vitaly Shmatikov. "Privacy-Preserving Deep Learning". In: Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security. CCS '15. Denver, Colorado, USA: ACM, 2015, pp. 1310–1321. isbn: 978-1-4503-3832-5. doi: 10.1145/2810103.2813687. url: http://doi.acm.org/10.1145/2810103.2813687.

[30] Craig Gentry and Shai Halevi. "Implementing Gentry's fully-homomorphic encryption scheme". In: Advances in Cryptology (EUROCRYPT). 2010.

[31] Ahmed El-Yahyaoui and Mohamed Dafir Elkettani. "Fully homomorphic encryption: state of art and comparison". In: International Journal of Computer Science and Information Security 14.4 (2016), p. 159.

[32] Aviad Kipnis and Eliphaz Hibshoosh. "Efficient Methods for Practical Fully Homomorphic Symmetric-key Encrypton, Randomization and Verification". In: IACR Cryptology ePrint Archive (2012), p. 637.

[33] Ran Gilad-Bachrach, Nathan Dowlin, Kim Laine, Kristin E. Lauter, Michael Naehrig, and John Robert Wernsing. "CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy". In: International Conference on Machine Learning (ICML). 2016.

[34] Ehsan Hesamifard, Hassan Takabi, and Mehdi Ghasemi. "CryptoDL: Deep Neural Networks over Encrypted Data". In: ArXiv abs/1711.05189 (2017).

[35] Hervé Chabanne, Amaury de Wargny, Jonathan Milgram, Constance Morel, and Emmanuel Prouff. "Privacy-Preserving Classification on Deep Neural Network". In: IACR Cryptology ePrint Archive 2017 (2017), p. 35.

[36] Yann Lecun, Leon Bottou, Y. Bengio, and Patrick Haffner. "Gradient-Based Learning Applied to Document Recognition". In: Proceedings of the IEEE 86 (Dec. 1998), pp. 2278–2324. doi: 10.1109/5.726791.

[37] Viorel Mihalef, Lucian Itu, Tommaso Mansi, and Puneet Sharma. "Lumped Parameter Whole Body Circulation Modelling". In: Patient-specific Hemodynamic Computations: Application to Personalized Diagnosis of Cardiovascular Pathologies. Ed. by Lucian Mihai Itu, Puneet Sharma, and Constantin Suciu. Cham: Springer International Publishing, 2017, pp. 111–152.

[38] Lucian Mihai Itu, Puneet Sharma, Bogdan Georgescu, Ali Kamen, Constantin Suciu, and Dorin Comaniciu. "Model based non-invasive estimation of PV loop from echocardiography". In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2014), pp. 6774–6777.

[39] Thomas James Ryan. "The coronary angiogram and its seminal contributions to cardiovascular medicine over five decades." In: Circulation 106 6 (2002), pp. 752–6.

[40] Vivian G. Ng and Alexandra J Lansky. "Novel QCA methodologies and angiographic scores". In: The International Journal of Cardiovascular Imaging 27 (2010), pp. 157–165.

[41] Armin Arbab-Zadeh. "What Will it Take to Retire Invasive Coronary Angiography". In: JACC. Cardiovascular imaging 9 5 (2016), pp. 565–7.

[42] Monique Tröbs et al. "Comparison of Fractional Flow Reserve Based on Computational Fluid Dynamics Modeling Using Coronary Angiographic Vessel Morphology Versus Invasively Measured Fractional Flow Reserve." In: The American journal of cardiology 117 1 (2016), pp. 29–35.

[43] Lucian Mihai Itu, Saikiran Rapaka, Tiziano Passerini, Bogdan Georgescu, Chris Schwemmer, Max Schoebinger, Thomas G. Flohr, Puneet Sharma, and Dorin Comaniciu. "A machine-learning approach for computation of fractional flow reserve from coronary computed tomography." In: Journal of applied physiology 121 1 (2016), pp. 42–52.

Abstract

In recent years, powered by state-of-the-art achievements in a broad range of areas, machine learning, with emphasis on deep neural networks, has received considerable attention from the healthcare sector.

The present work focuses on the exploration, development, and evaluation of deep learning-based solutions for automatic medical data analysis. The final goal of the current thesis is to incorporate learning-based solutions in the medical imaging analysis pipeline, starting from image acquisition, continuing with image interpretation and up to secure patient health data manipulation.

To exploit the potential of deep learning-based methods for medical imaging analysis, an advanced type of breast imaging, three-dimensional mammography has been considered. Moreover, the reconstruction, which represents the crucial component in producing images of the internal structure of the human body, has been tackled by integrating deep-learning with physics and acquisition geometry of computed tomography (CT). To allow for the deep-learning-based analysis to be performed on medical data without disclosing patient-related health information, privacy-preserving deep learning solutions that operate directly on homomorphically encrypted data have been proposed.

While still far from being deemed trustworthy solutions for practical medical image analysis, the results hence obtained reflect the potential of learning-based approaches to shaping the future of the healthcare industry.

În ultimii ani, ca urmare a avansului tehnologic, popularitatea inteligenței artificiale a explodat afectând un spectru larg al domeniilor de activitate, printre care și sectorul medical.

Lucrarea de față se concentrează pe exploatarea, dezvoltarea și evaluarea soluțiilor bazate pe rețele neurale adânci (eng. Deep Learning) pentru analiza automată a datelor medicale. Astfel, scopul final al prezentei teze este de a include soluții bazate pe învățare în procesul de analiză și prelucrare a imaginilor medicale, pornind de la achiziția imaginii, continuând cu interpretarea acesteia și până la asigurarea confidențialității datelor cu caracter personal în vederea manipulării acestora.

În acest context, pentru a exploata potențialul modelelor neurale adânci în analiza imagisticii medicale, s-a propus o metodă de detecție automată a maselor de țesut tumoral la nivelul sânului din mamografii digitale cu tomosinteză (mamografia 3D). Pentru a demonstra avantajele pe care rețelele neurale adânci le pot aduce în procesul de formare al imaginilor medicale, reconstrucția, care reprezintă componenta esențială în reprezentarea structurii interne a corpului uman, este adresată prin combinarea rețelelor neurale adânci cu noțiunile de bază ale fizicii și geometriei ce descriu sistemul de achiziție al tomografiei computerizate (CT). În finalul tezei, s-a propus o soluție care permite realizarea unor aplicații de medicină personalizată bazate pe inteligență artificială și care, în același timp, protejează datele personale ale pacienților. Soluția propusă se bazează pe rețele neurale adânci și pe criptarea homomorfică, un tip special de criptare care permite realizarea de operații, de exemplu aritmetice, asupra informațiilor criptate.

Deși încă nu au ajuns la nivelul de încredere necesar pentru a putea fi utilizate în practica medicală, rezultatele obținute reflectă potențialul soluțiilor bazate pe rețele neurale adânci în conturarea viitorul sistemului medical.