Interpreting the output of dose maps predicted by a Unet with the γ-index analysis

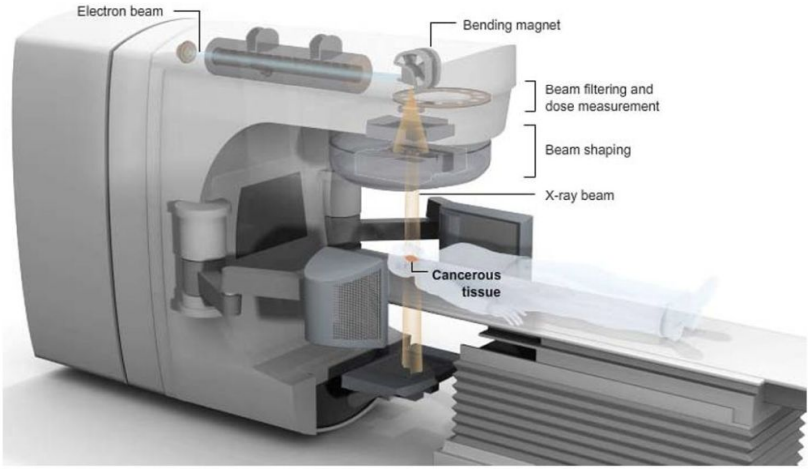
next-AIM workshop on XAI techniques for medical data analysis

lorenzo.marini@phd.unipi.it

16th-18th October 2024, Bari

What is Radiotherapy?

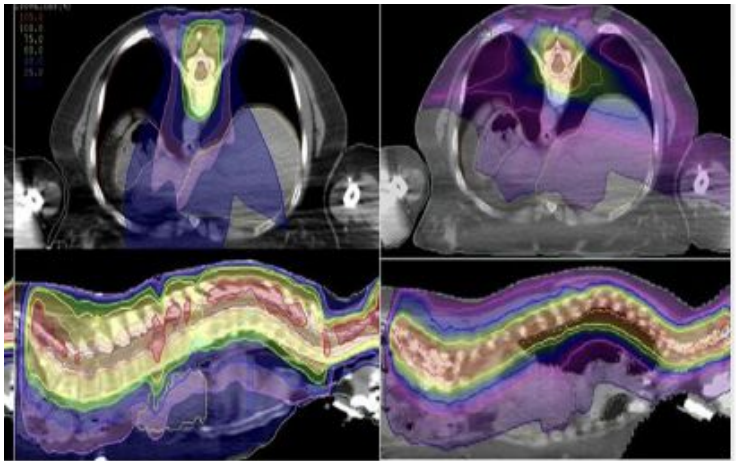
RT consists of the medical **treatment of tumors**, by exposing the affected tissues to **ionizing radiation** (e.g., X-rays, e-)



RT has advanced, but it requires sophisticated systems to **ensure correct dose administration**.

Errors caused by **machinery malfunctioning**, in patient **positioning** or changes in **anatomy** must be detected to prevent accidents and treatment errors.

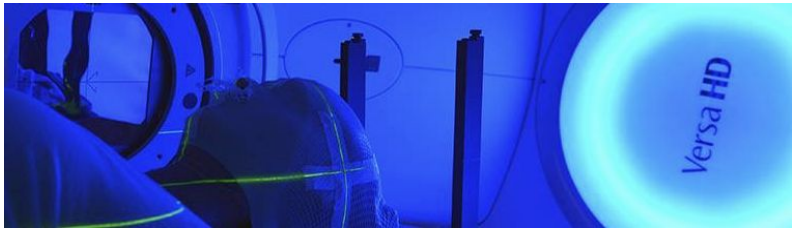
IMRT VMAT



Major accidents in RT

Equipment malfunction

- Accelerator software problems (USA and Canada)
- Incorrect repair of accelerator (Spain)
- Accelerator interlock failure (Poland)



Calibration of RT equipment

- Incorrect decay data (USA)
- Miscalibration of beam (Costa Rica)

Errors in treatment planning

- Erroneous use of TPS (UK)
- Computer file not updated (USA)
- Errors in TPS data entry (Panama)

IAEA
International Atomic Energy Agency

Better technologies are needed to **verify treatment quality.**

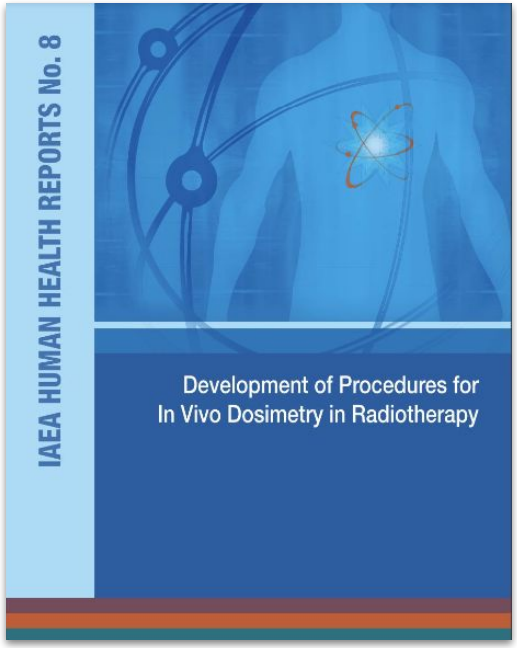
In-Vivo Dosimetry: a possible solution

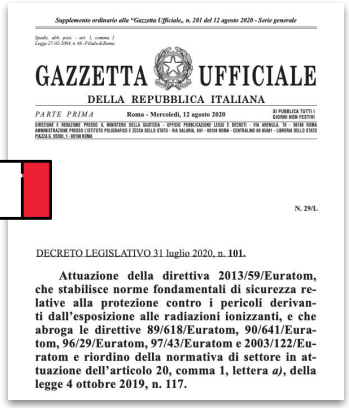
IVD is a **direct method for measuring radiation doses** in cancer patients undergoing RT.

Purpose: ensure treatment is executed as prescribed.

Safety & Compliance:

- Acts as a safety measure for dose delivery.
- Meets patient **radiation protection standards** set by **national regulatory bodies**.





Supplemento ordinario alla "Gazzetta Ufficiale", n. 201 del 12 agosto 2020 - Serie generale

GAZZETTA UFFICIALE

DELLA REPUBBLICA ITALIANA

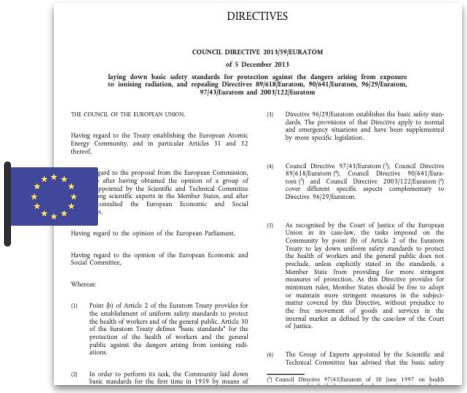
PARTE PRIMA Roma - Mercoledì, 12 agosto 2020

N. 29/L

DECRETO LEGISLATIVO 31 luglio 2020, n. 101.

Attuazione della direttiva 2013/59/Euratom, che stabilisce norme fondamentali di sicurezza relative alla protezione contro i pericoli derivanti dall'esposizione alle radiazioni ionizzanti, e che abroga le direttive 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom e 2003/122/Euratom e riordino della normativa di settore in attuazione dell'articolo 20, comma 1, lettera *a)*, della legge 4 ottobre 2019, n. 117.

[DECRETO LEGISLATIVO 31 luglio 2020, n. 101]



DIRECTIVES

COUNCIL DIRECTIVE 2013/59/EURATOM

of 5 December 2013

laying down basic safety standards for protection against the dangers arising from exposure to ionising radiation, and repealing Directives 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom and 2003/122/Euratom

THE COUNCIL OF THE EUROPEAN UNION,

Having regard to the Treaty establishing the European Atomic Energy Community, and in particular Articles 31 and 32 thereof,

Having regard to the opinion of the European Parliament,

Having regard to the opinion of the European Economic and Social Committee,

Whereas:

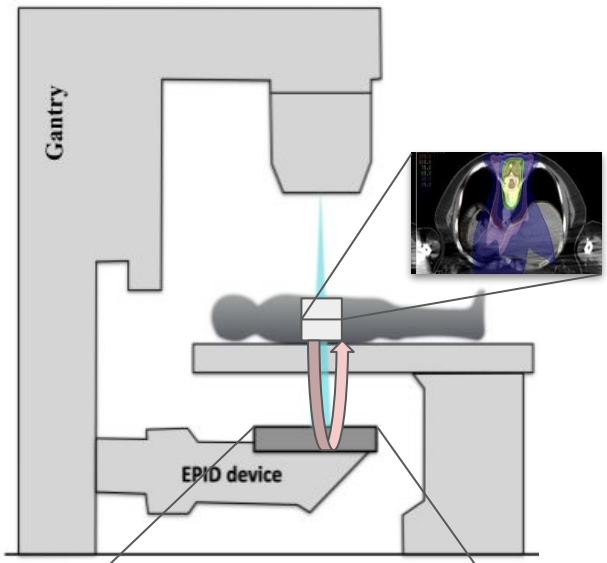
[COUNCIL DIRECTIVE 2013/59/EURATOM]

In vivo dosimetry involves the measurement of radiation doses to patients during their radiation treatment in order to ensure that the treatments are carried out as they were intended. For many years, it has been common practice to use in vivo dosimetry to check doses to organs at risk (e.g. skin, eye or rectum). The primary goal of in vivo dosimetry, however, is quality assurance (QA) of the radiotherapy process. It is considered an important part of quality management of a radiotherapy department. Following recommendations by the World Health Organization (WHO), the International Commission on Radiological Protection (ICRP), the IAEA [1–3] and other bodies [4–6], the use of in vivo dosimetry has become more widespread.

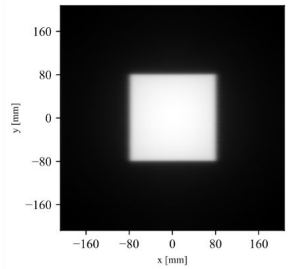
In vivo dosimetry is used for the overall verification of the chain of treatment preparation and delivery. As such, it measures the radiation dose to the patient, which can be affected by many variables in the overall radiotherapy process. The global results of measurements of patient doses provide the information necessary for assessment of the accuracy and precision in dose planning and delivery for a specific treatment site, or by a given radiotherapy machine. In vivo dosimetry can also be used for the estimation of uncertainties in radiation treatment at a given institution.

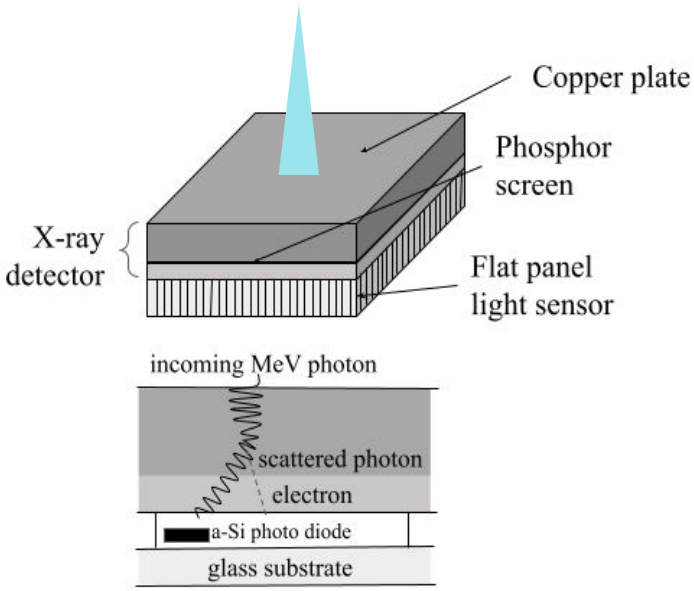
[IAEA]

Electronic Portal Imaging Device for IVD



EPIDs detectors capture **X-ray fluence** on a pixel-based surface to produce **2D digital images**





Limitations

Current methods are based on **physical models** and MC simulations:

- **time-consuming**
- **complex**
- limited **clinical applicability**

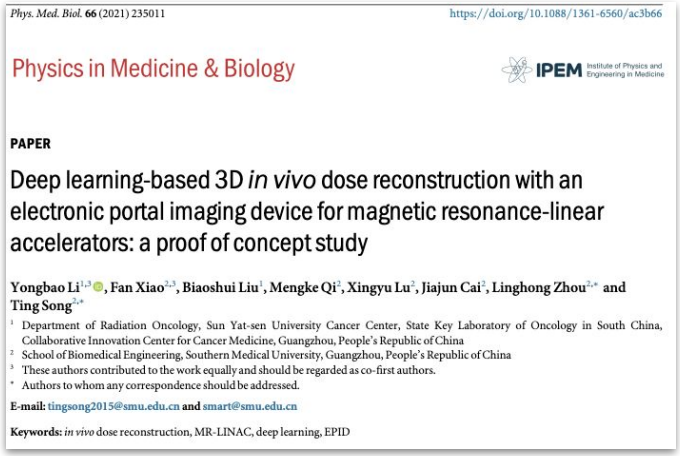
Advantages

1) High **resolution**
2) **rapid** image capture
3) long-term **stability**

Opportunity for DL

Use DL models to **simplify EPID dose reconstruction** without traditional physical models.

Can AI be explored to address this issues?



Phys. Med. Biol. 66 (2021) 235011 https://doi.org/10.1088/1361-6560/ac3b66

Physics in Medicine & Biology — IPEM Institute of Physics and Engineering in Medicine

PAPER

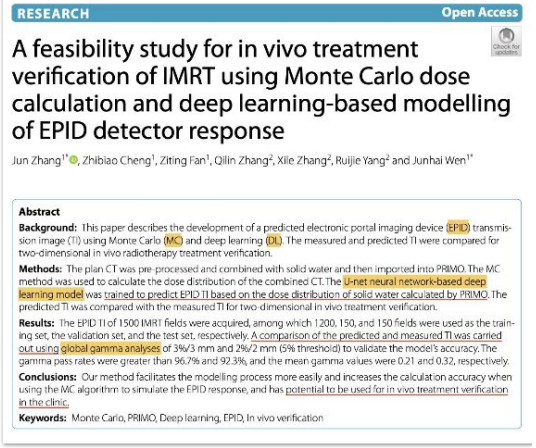
Deep learning-based 3D *in vivo* dose reconstruction with an electronic portal imaging device for magnetic resonance-linear accelerators: a proof of concept study

Yongbao Li[1,3], Fan Xiao[2,3], Biaoshui Liu[1], Mengke Qi[2], Xingyu Lu[2], Jiajun Cai[2], Linghong Zhou[2,*] and Ting Song[2,*]

[1] Department of Radiation Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, People's Republic of China
[2] School of Biomedical Engineering, Southern Medical University, Guangzhou, People's Republic of China
[3] These authors contributed to the work equally and should be regarded as co-first authors.
[*] Authors to whom any correspondence should be addressed.

E-mail: tingsong2015@smu.edu.cn and smart@smu.edu.cn

Keywords: *in vivo* dose reconstruction, MR-LINAC, deep learning, EPID



RESEARCH Open Access

A feasibility study for in vivo treatment verification of IMRT using Monte Carlo dose calculation and deep learning-based modelling of EPID detector response

Jun Zhang[1*], Zhibiao Cheng[1], Ziting Fan[1], Qilin Zhang[2], Xilie Zhang[2], Ruijie Yang[2] and Junhai Wen[1*]

Abstract

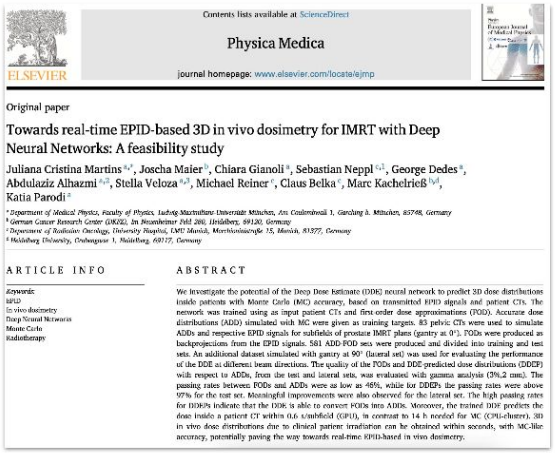
Background: This paper describes the development of a predicted electronic portal imaging device (EPID) transmission image (TI) using Monte Carlo (MC) and deep learning (DL). The measured and predicted TI were compared for two-dimensional in vivo radiotherapy treatment verification.

Methods: The plan CT was pre-processed and combined with solid water and then imported into PRIMO. The MC method was used to calculate the dose distribution of the combined CT. The U-net neural network-based deep learning model was trained to predict EPID TI based on the dose distribution of solid water calculated by PRIMO. The predicted TI was compared with the measured TI for two-dimensional in vivo treatment verification.

Results: The EPID TI of 1500 IMRT fields were acquired, among which 1200, 150, and 150 fields were used as the training set, the validation set, and the test set, respectively. A comparison of the predicted and measured TI was carried out using global gamma analyses of 3%/3 mm and 2%/2 mm (5% threshold) to validate the model's accuracy. The gamma pass rates were greater than 96.7% and 92.3%, and the mean gamma values were 0.21 and 0.32, respectively.

Conclusions: Our method facilitates the modelling process more easily and increases the calculation accuracy when using the MC algorithm to simulate the EPID response, and has potential to be used for in vivo treatment verification in the clinic.

Keywords: Monte Carlo, PRIMO, Deep learning, EPID, In vivo verification



Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

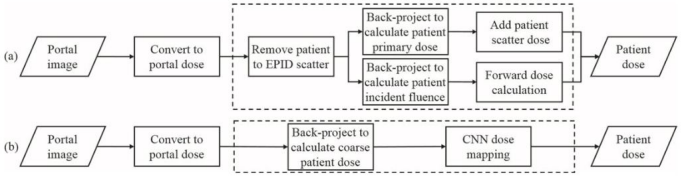
Original paper

Towards real-time EPID-based 3D in vivo dosimetry for IMRT with Deep Neural Networks: A feasibility study

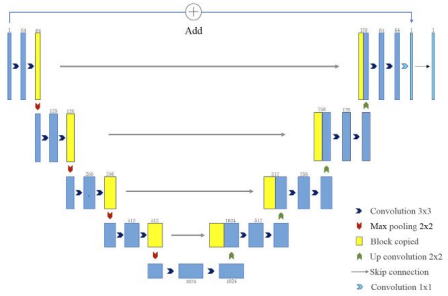
Juliana Cristina Martins[a,*], Joscha Maier[b], Chiara Gianoli[a], Sebastian Neppl[c,d], George Dedes[a], Abdulaziz Alhazmi[c,d], Stella Veloza[a,b], Michael Reiner[c], Claus Belka[c], Marc Kachelrieß[b,e], Katia Parodi[a]

ARTICLE INFO

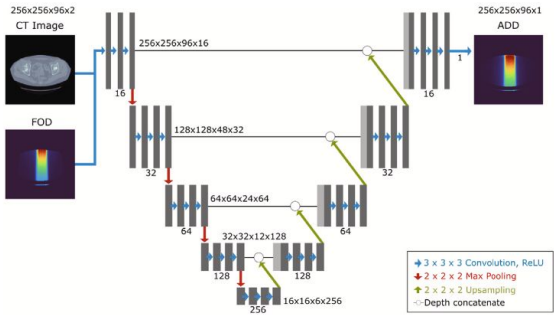
ABSTRACT



https://doi.org/10.1088/1361-6560/ac3b66



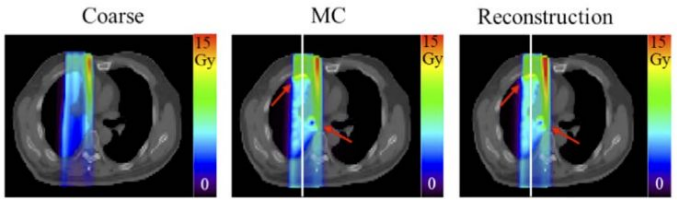
https://doi.org/10.1186/s13014-022-01999-3

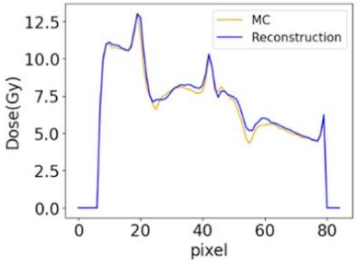


https://www.physicamedica.com/article/S1120-1797(23)00175-8/abstract

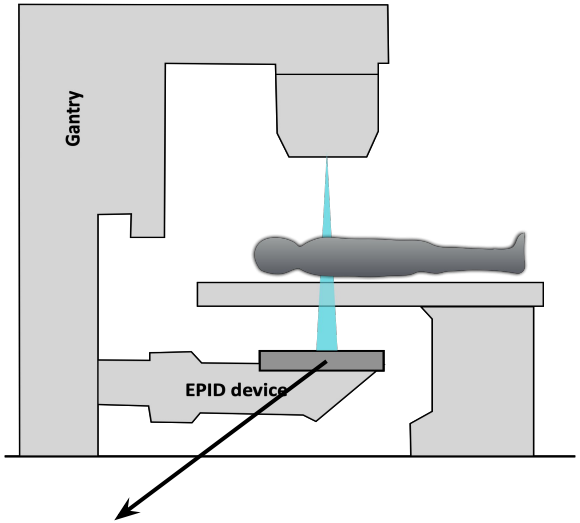
STATE-OF-THE-ART

- only **few publications**
- mainly on **simulated data**
- only **U-net** architecture explored
- only one **loss function** for training phase explored (MSE)

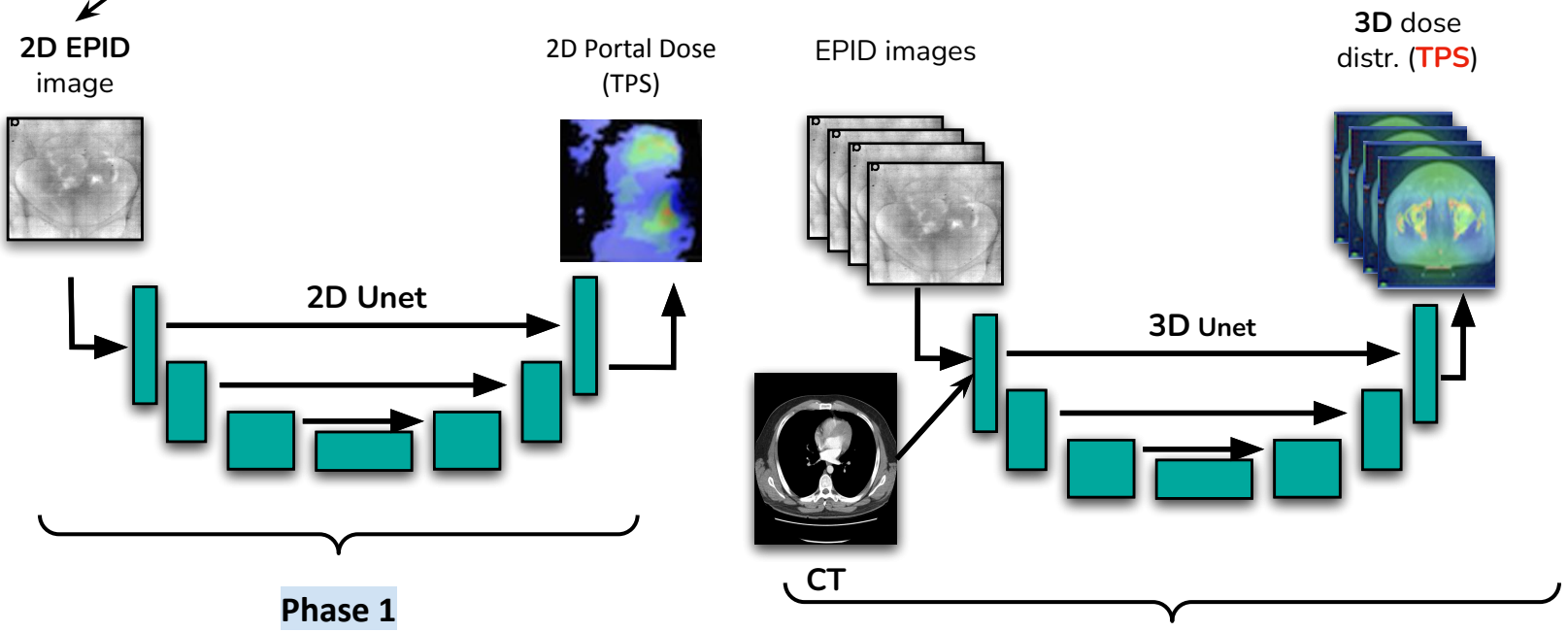




Goal of the Ph.D. project



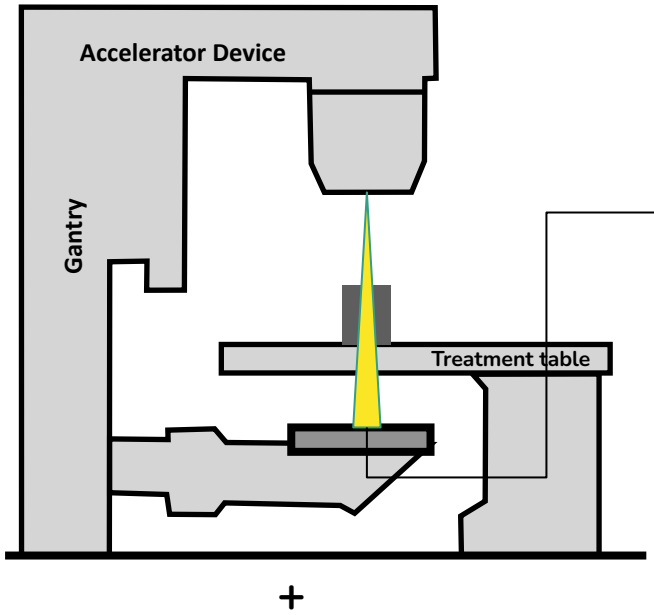
The main purpose of this Ph.D. project is to develop a **multi-input DL-based 3D in-vivo dose reconstruction framework** based on **real EPID images** acquired during dose delivery.



Data exploration

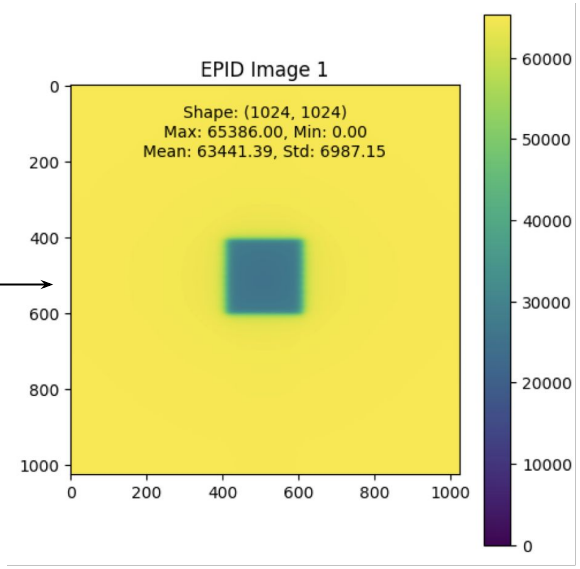
We collected **210 pairs** of EPID-PD images from various phantoms representing different material densities (lung, solid water, titanium, and bone), along with corresponding simulated dose images).

Experimental Setup

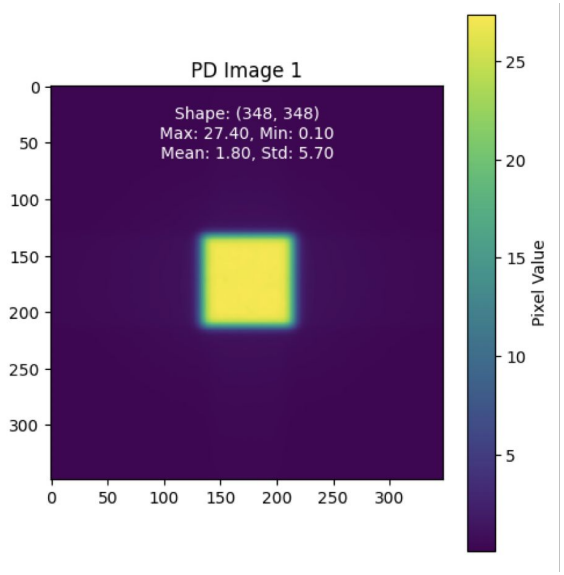


+

MONACO Elekta → Treatment Planning System

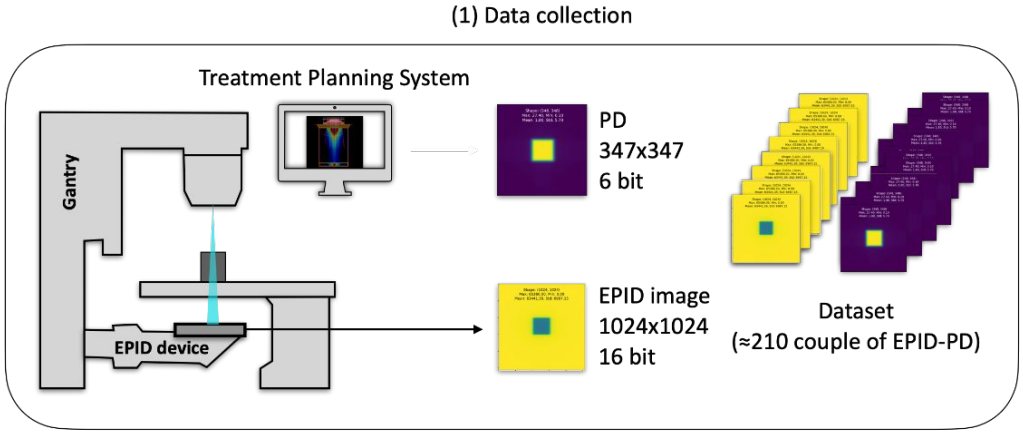


- Size: 1024x1024
- 16 bit
- Pixel spacing: [0.405, 0.405] mm
- **Grey-scale** pixel value

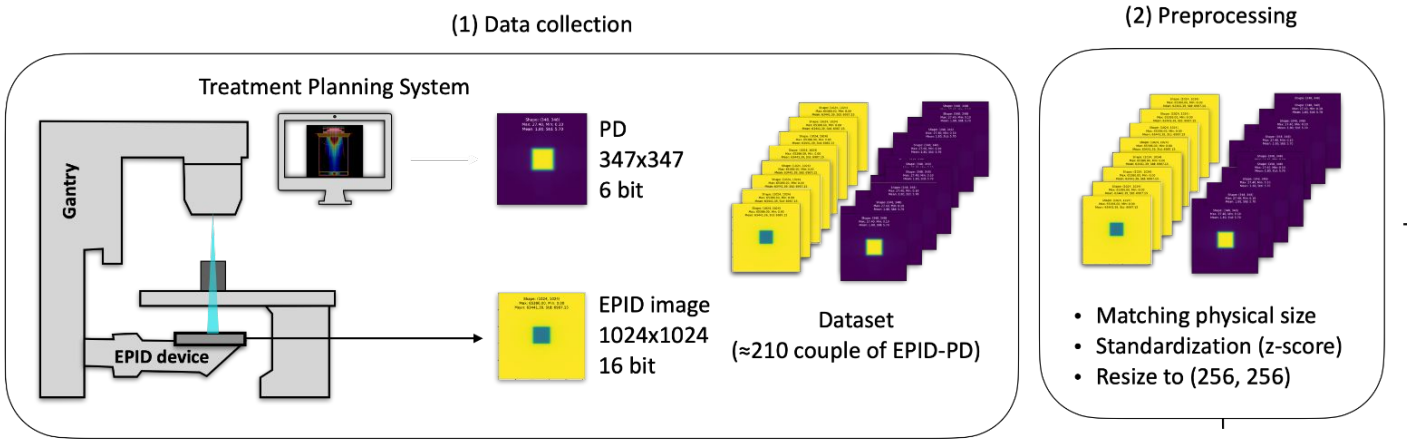


- Size: 347x347
- Range: [0, ≈54]
- Pixel spacing: [1.0, 1.0] mm
- **Gray dose** pixel value [**cGy**]

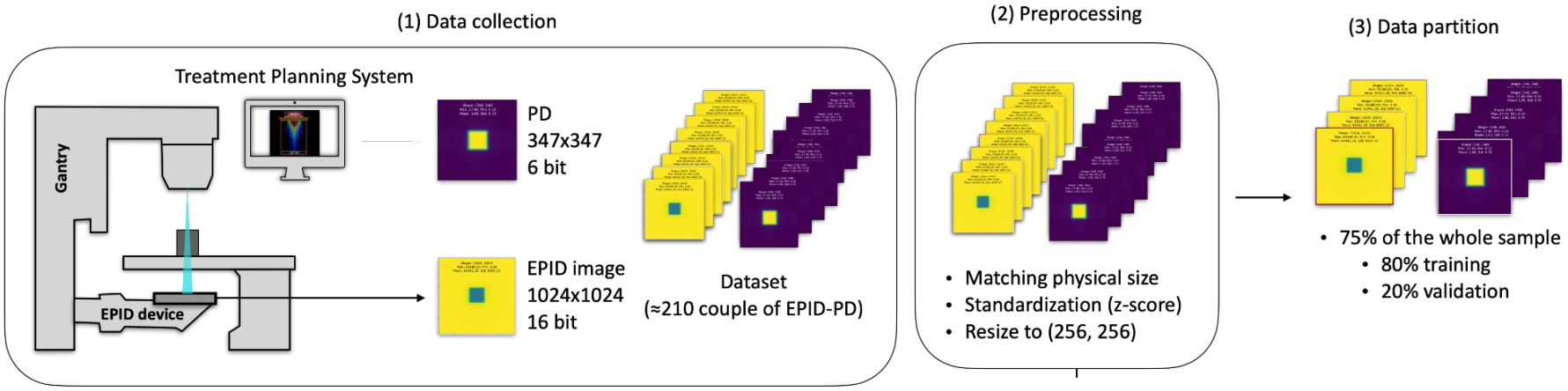
Pipeline of the analysis



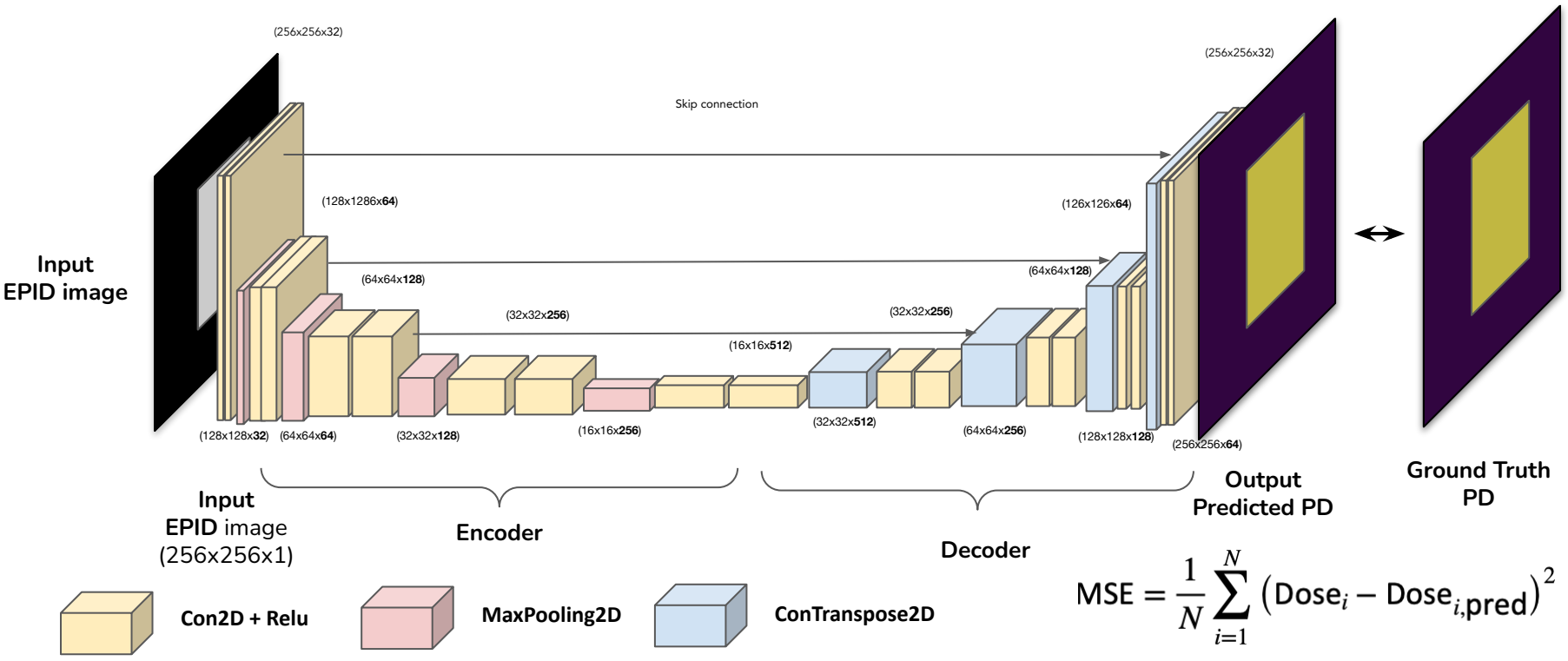
Pipeline of the analysis



Pipeline of the analysis



Unet architecture



$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\text{Dose}_i - \text{Dose}_{i,\text{pred}}\right)^2$$

Model Architecture

U-net designed to map EPID images to PD distributions.

Optimization

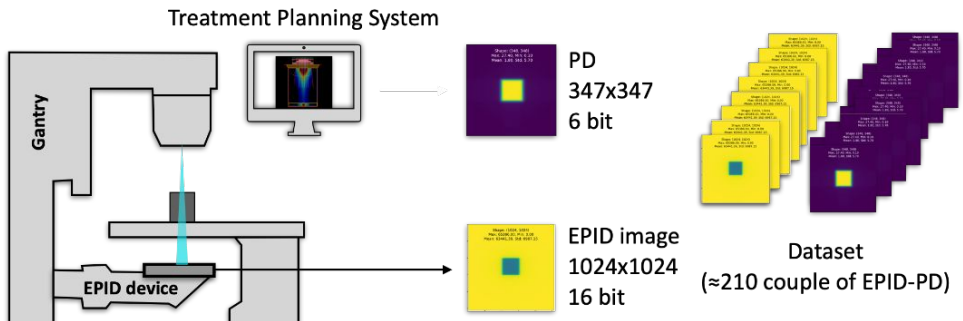
Loss function: Mean Square Error (MSE) | Optimizer: Adam | Batch size of 8, 200 epochs.

Evaluation Metric

Mean Absolute Error (MAE) between model predictions and true PD images.

Pipeline of the analysis

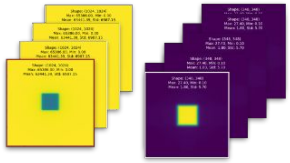
(1) Data collection



(2) Preprocessing

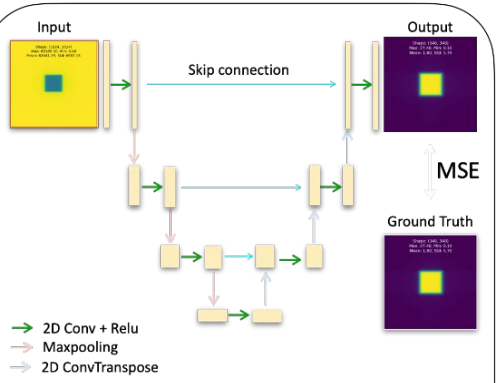
- Matching physical size
- Standardization (z-score)
- Resize to (256, 256)

(3) Data partition



- 75% of the whole sample
  - 80% training
  - 20% validation

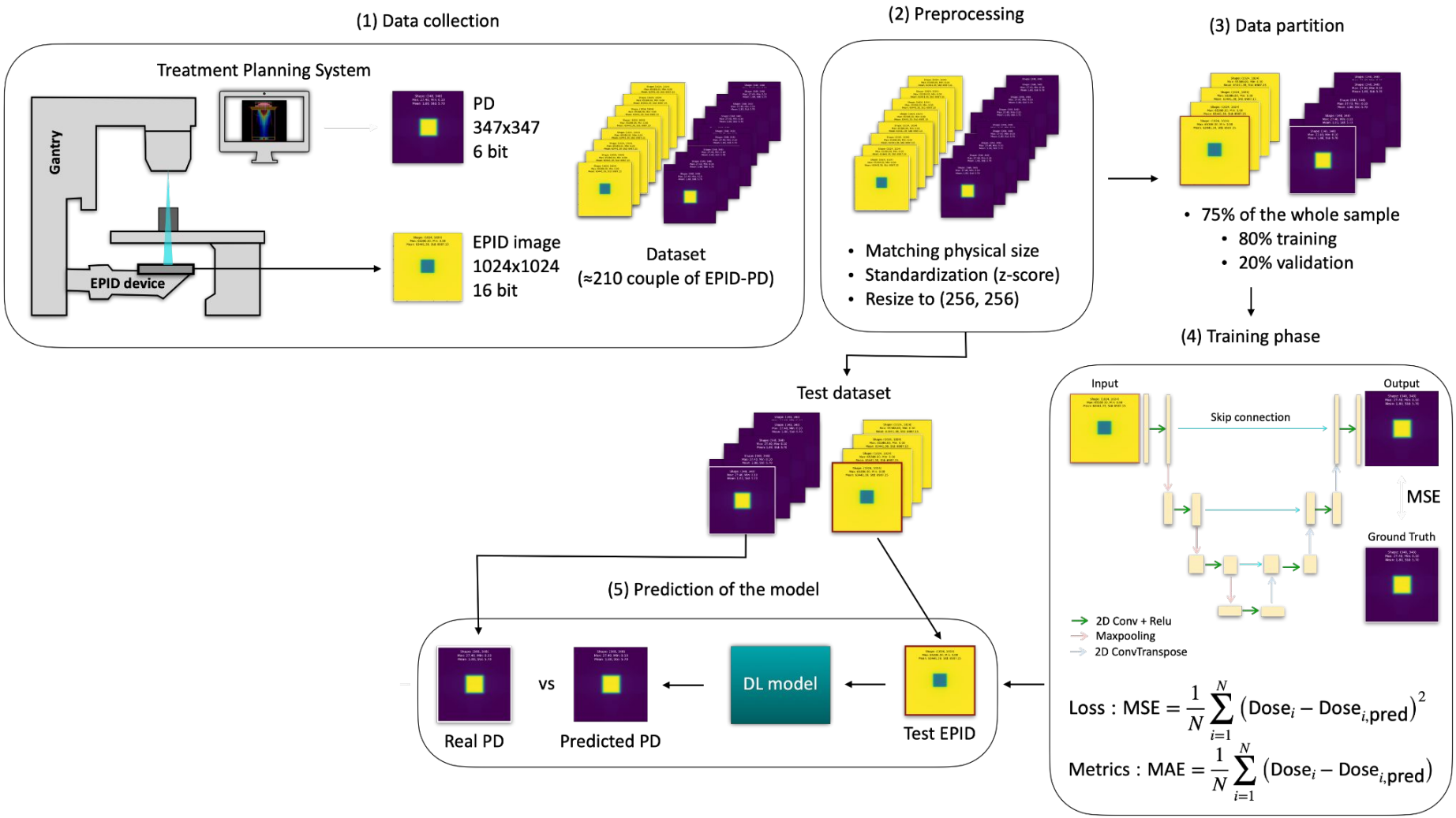
(4) Training phase



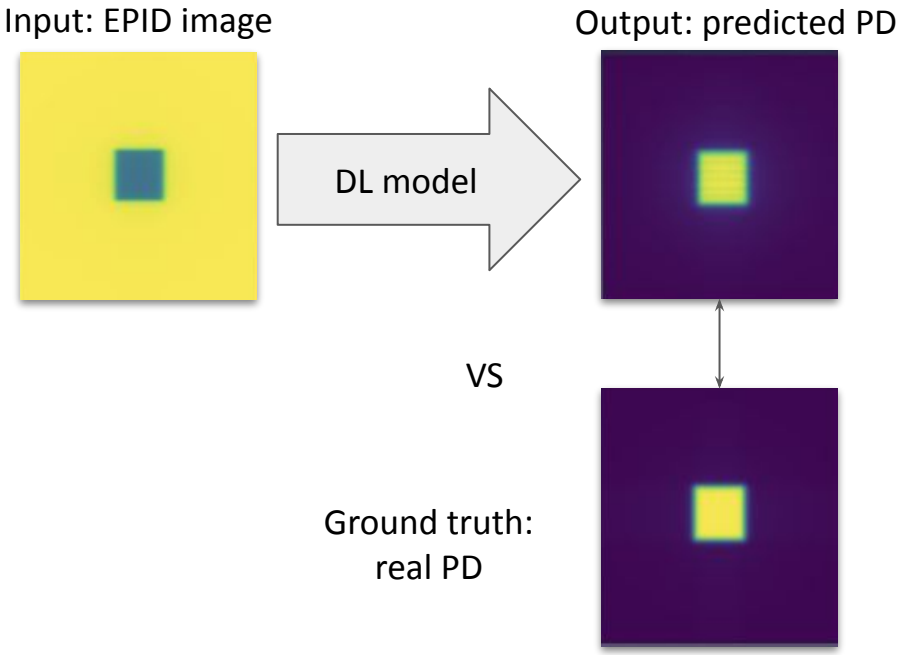
$$\text{Loss : MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\text{Dose}_i - \text{Dose}_{i,\text{pred}}\right)^2$$

$$\text{Metrics : MAE} = \frac{1}{N}\sum_{i=1}^{N}\left(\text{Dose}_i - \text{Dose}_{i,\text{pred}}\right)$$

Pipeline of the analysis



How to evaluate the accuracy of the reconstruction?



γ-index analysis

- The agreement between simulated and reconstructed doses is assessed by **comparing both dose and position**.
- The **γ-index** merges **dose difference** and **distance to agreement** (DTA) into one metric.
- Points with **γ > 1 fail the test**.
- A plane is **clinically acceptable** if over **95% of points have γ ≤ 1**.

$$\Gamma(\mathbf{r}_{real}, \mathbf{r}_{pred}) = \sqrt{\frac{\Delta \mathbf{r}^2(\mathbf{r}_{real}, \mathbf{r}_{pred})}{\delta r^2} + \frac{\Delta D^2(\mathbf{r}_{real}, \mathbf{r}_{pred})}{\delta D^2}}$$

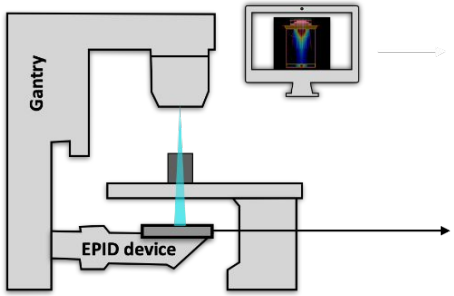
$$\gamma(\mathbf{r}_{real}, \mathbf{r}_{pred}) = \min\{\Gamma(\mathbf{r}_{real}, \mathbf{r}_{pred})\} \; \forall \mathbf{r}_{real}$$

$$\gamma\text{-passing rate} = \frac{\#\gamma < 1}{\#\text{tot}\gamma}$$

Pipeline of the analysis

(1) Data collection

Treatment Planning System



PD
347x347
6 bit

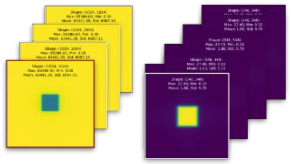
EPID image
1024x1024
16 bit

Dataset
(≈210 couple of EPID-PD)

(2) Preprocessing

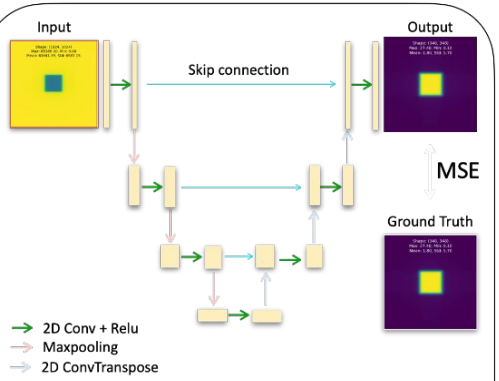
- Matching physical size
- Standardization (z-score)
- Resize to (256, 256)

(3) Data partition



- 75% of the whole sample
  - 80% training
  - 20% validation

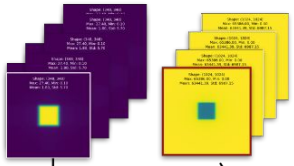
(4) Training phase



$$\text{Loss : MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\text{Dose}_i - \text{Dose}_{i,\text{pred}}\right)^2$$

$$\text{Metrics : MAE} = \frac{1}{N}\sum_{i=1}^{N}\left(\text{Dose}_i - \text{Dose}_{i,\text{pred}}\right)$$

Test dataset



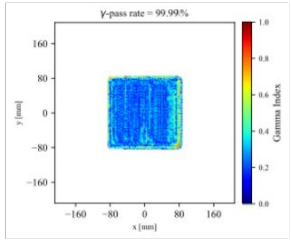
(5) Prediction of the model

(6) Evaluation of the performance

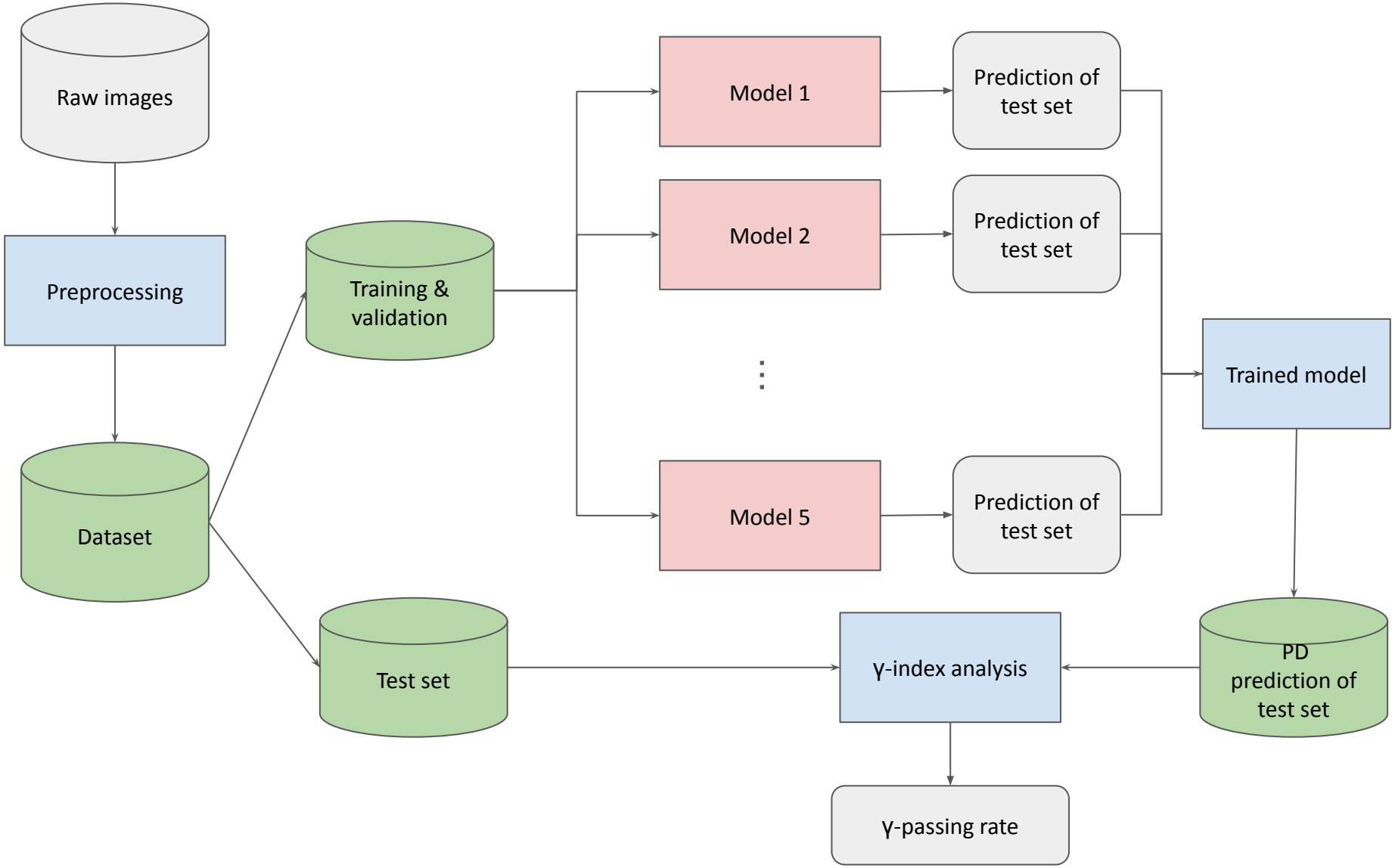
$$\Gamma(\mathbf{r}_{\text{real}}, \mathbf{r}_{\text{pred}}) = \sqrt{\frac{\Delta r^2(\mathbf{r}_{\text{real}}, \mathbf{r}_{\text{pred}})}{\delta r^2} + \frac{\Delta D^2(\mathbf{r}_{\text{real}}, \mathbf{r}_{\text{pred}})}{\delta D^2}}$$

$$\gamma(\mathbf{r}_{\text{real}}, \mathbf{r}_{\text{pred}}) = \min\{\Gamma(\mathbf{r}_{\text{real}}, \mathbf{r}_{\text{pred}})\}\,\forall\,\mathbf{r}_{\text{real}}$$

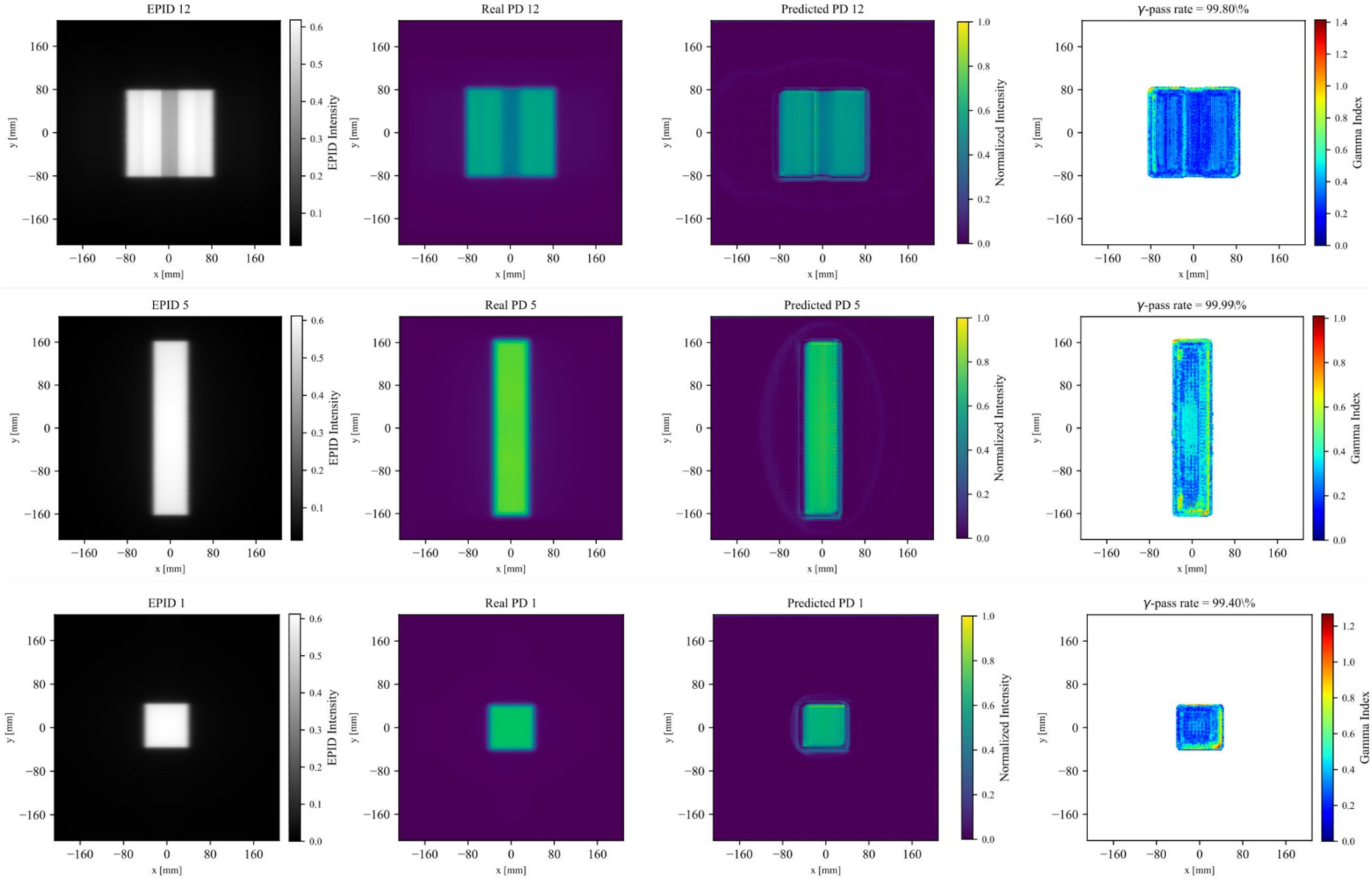
$$\gamma\text{-pass rate} = \frac{\#\gamma < 1}{\#tot\gamma}$$



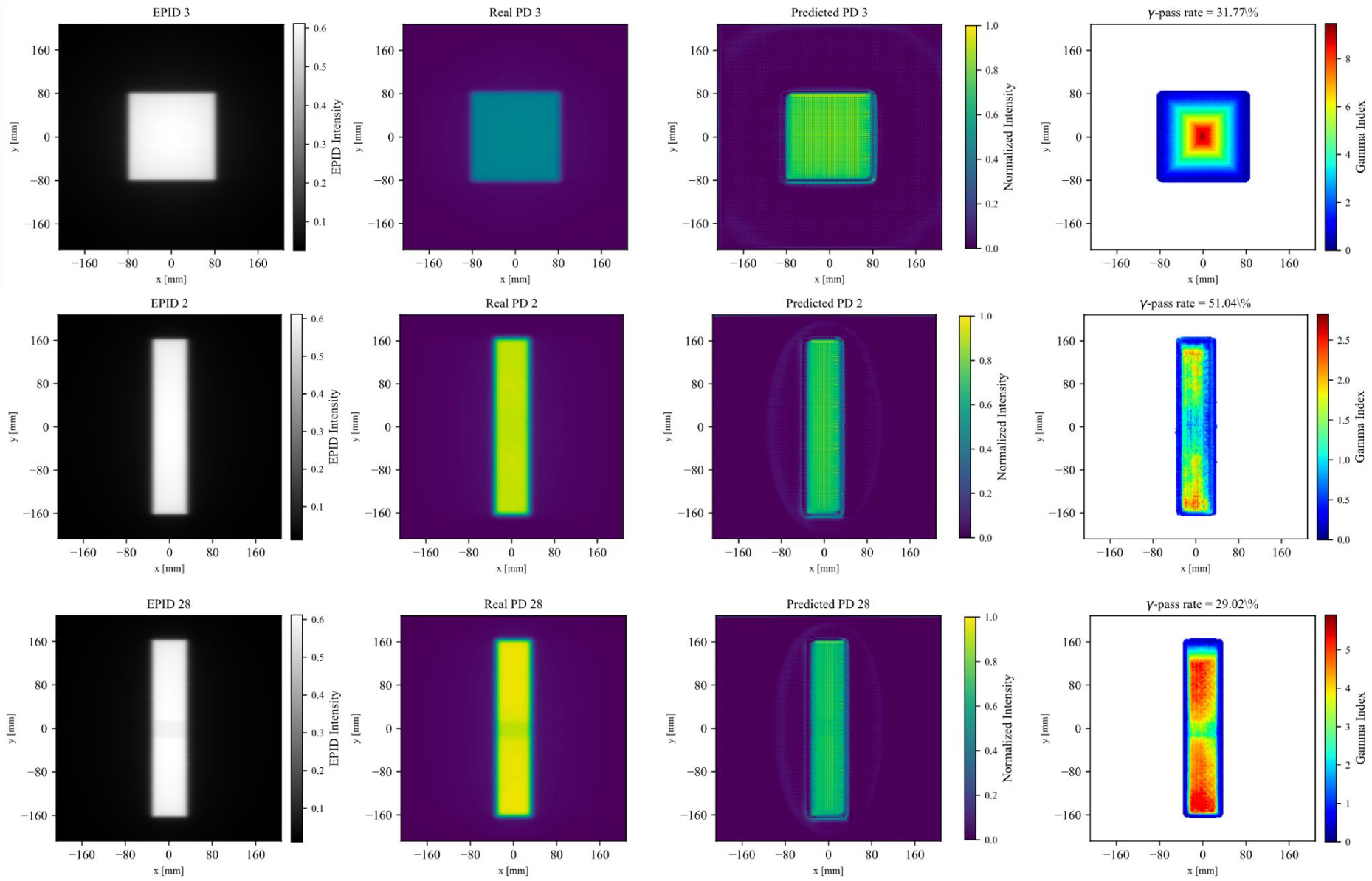
Pipeline of the analysis



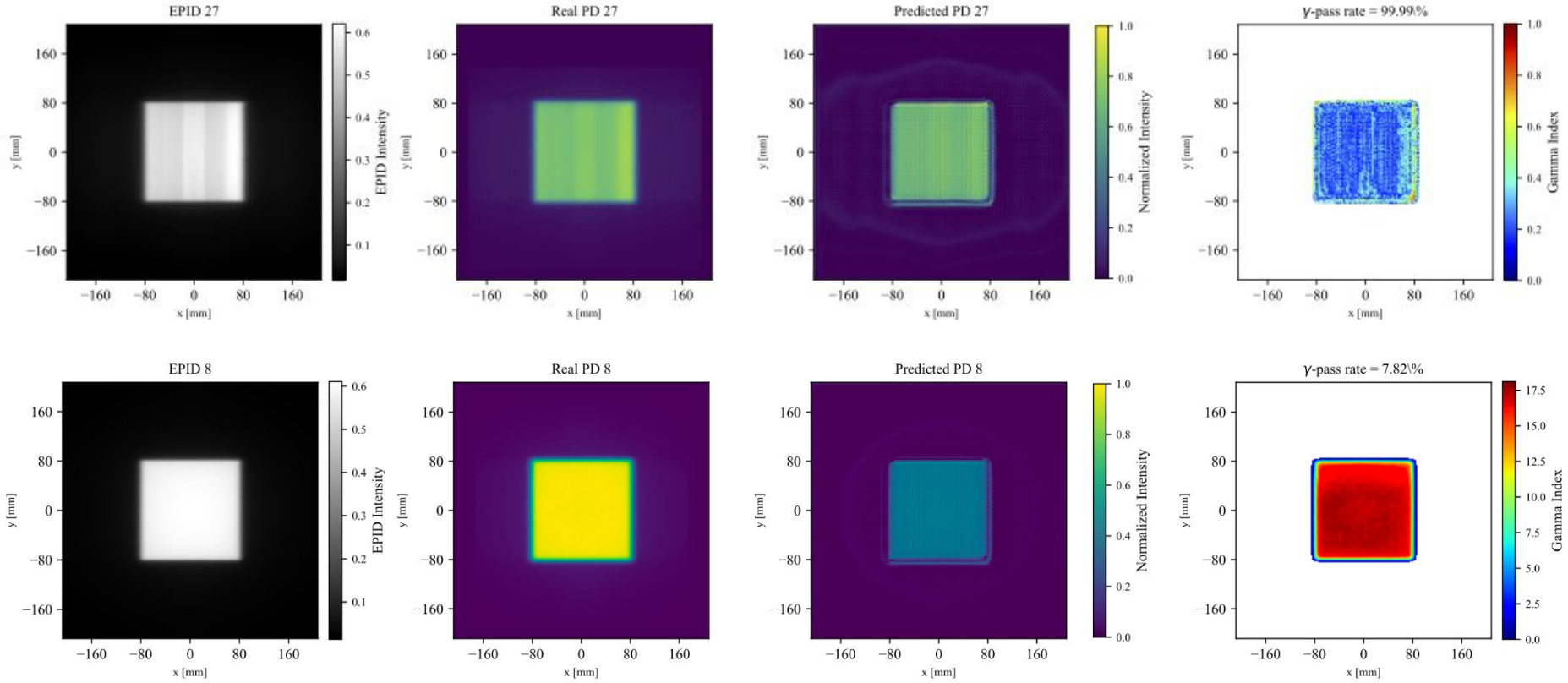
Preliminary results



Preliminary results



worse and best cases



Best-Case Scenarios
y-passing rate **≥99%** in some test samples.

Overall Performance
Average y-passing rate of **(81.50 ± 4.45)%** across all test samples.

The model predicts the PD in **under 1 second,** significantly **faster than TPS** (≈ 20-30 min)

Improvements
Dataset and DL model | **Preprocessing phase** (normalization)

Thank you for the attention!

Questions?

- **Contacts:**

lorenzo.marini@phd.unipi.it
lorenzo.marini@pi.infn.it

- **My office**

Largo Bruno Pontecorvo, 3/Building C, 56127 Pisa PI, Room 142 (first floor)

Acknowledgments

InTrEPID: In vivo 3D dosimetry in radiotherapy Treatments with EPID. Research partly funded by MUR PRIN 2022CWXR8K - CUP I53D23000520006.

ARTEMIS: Artificial Intelligence in RadioTherapy with EPID Monitoring System funded by INFN.

Backup slides

References

W. Elmpt et al.
"*A literature review of electronic portal imaging for radiotherapy dosimetry*"
2008

B. Mijnheer et al.
"In vivo dosimetry in external beam radiotherapy"
2013

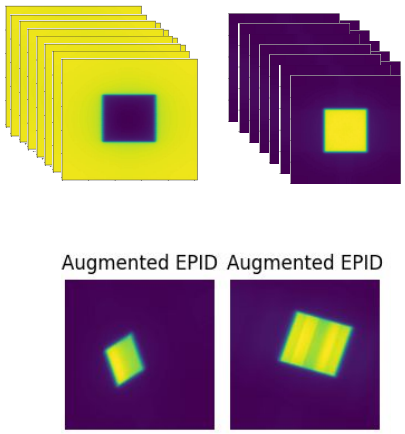
I. Olaciregui-Ruiz et al.
"*In vivo dosimetry in external beam photon radiotherapy: Requirements and future directions for research, development, and clinical practice*"
2020

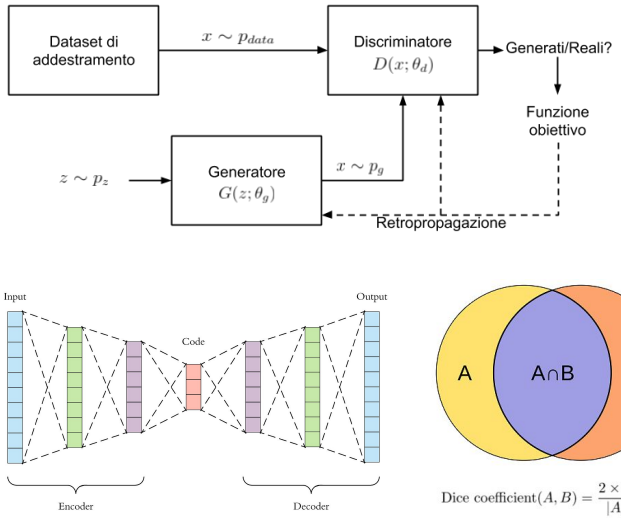
J. Martins et al.
"*Towards real-time EPID-based 3D in vivo dosimetry for IMRT with Deep Neural Networks: A feasibility study*"
2023

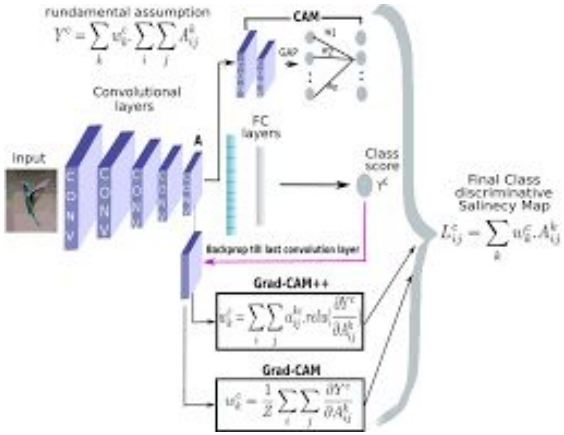
J. Zhang et al.
"*A feasibility study for in vivo treatment verification of IMRT using Monte Carlo dose calculation and deep learning-based modelling of EPID detector response*"
2022

F. Chan et al.
"*Integration of AI and Machine Learning in Radiotherapy QA*"
2020

Improvements







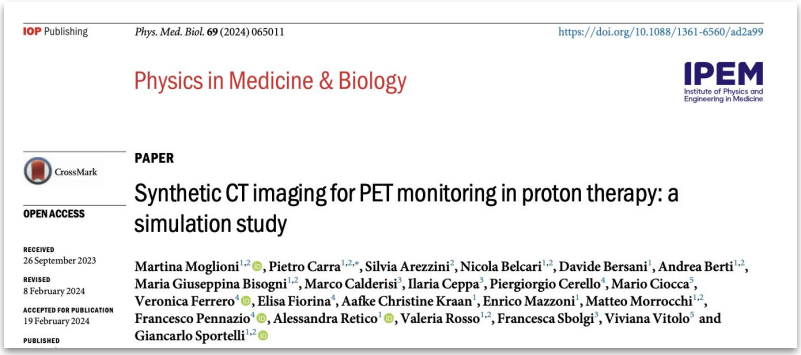
Dataset
Correction of artefacts
Data size
Heterogeneity
Data augmentation

DL model
Transfer learning
Different architectures
Custom loss function

Explainability
Transparency
Saliency map
GradCAM

Timeline of the project, and next steps

Phase	Task	2nd Year	2nd Year	3rd Year	3rd Year
I	**Data Generation**				
I.I	Data acquisition on phantoms				
I.II	TPS simulation				
II	**DL models**				
II.I	2D DL model				
II.II	3D DL model				
III	**PhD Thesis Compilation**				



IOP Publishing — Phys. Med. Biol. 69 (2024) 065011 — https://doi.org/10.1088/1361-6560/ad2a99

Physics in Medicine & Biology — IPEM Institute of Physics and Engineering in Medicine

CrossMark

PAPER

OPEN ACCESS

Synthetic CT imaging for PET monitoring in proton therapy: a simulation study

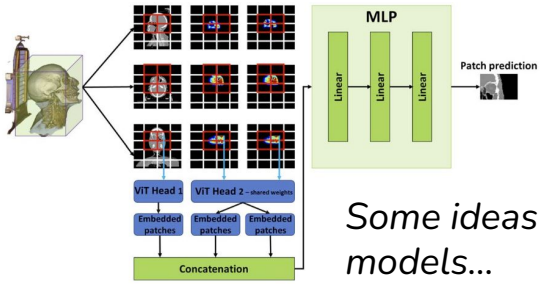
RECEIVED 26 September 2023

REVISED 8 February 2024

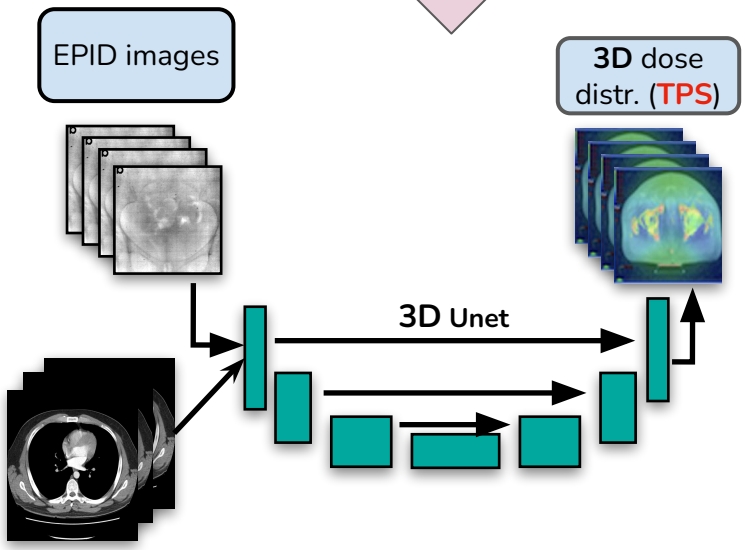
ACCEPTED FOR PUBLICATION 19 February 2024

PUBLISHED

Martina Moglioni[1,2], Pietro Carra[1,2,*], Silvia Arezzini[2], Nicola Belcari[1,2], Davide Bersani[1], Andrea Berti[1,2], Maria Giuseppina Bisogni[1,2], Marco Calderisi[3], Ilaria Ceppa[3], Piergiorgio Cerello[3], Mario Ciocca[5], Veronica Ferrero[4], Elisa Fiorina[4], Aafke Christine Kraan[1], Enrico Mazzoni[1], Matteo Morrocchi[1,2], Francesco Pennazio[4], Alessandra Retico[1], Valeria Rosso[1,2], Francesca Sbolgi[3], Viviana Vitolo[5] and Giancarlo Sportelli[1,2]



Some ideas regarding possible models...



Phase 2

Conclusions

Study Overview: We converted the measured EPID response into actual PD using a DL network and compared it with the simulated PD calculated by TPS.

Gamma-Analysis: A standard γ-analysis of **3%/3mm** was performed on the PD predicted by the DL network, with a mean γ-pass rate of **(81.50 ± 4.45)%**.

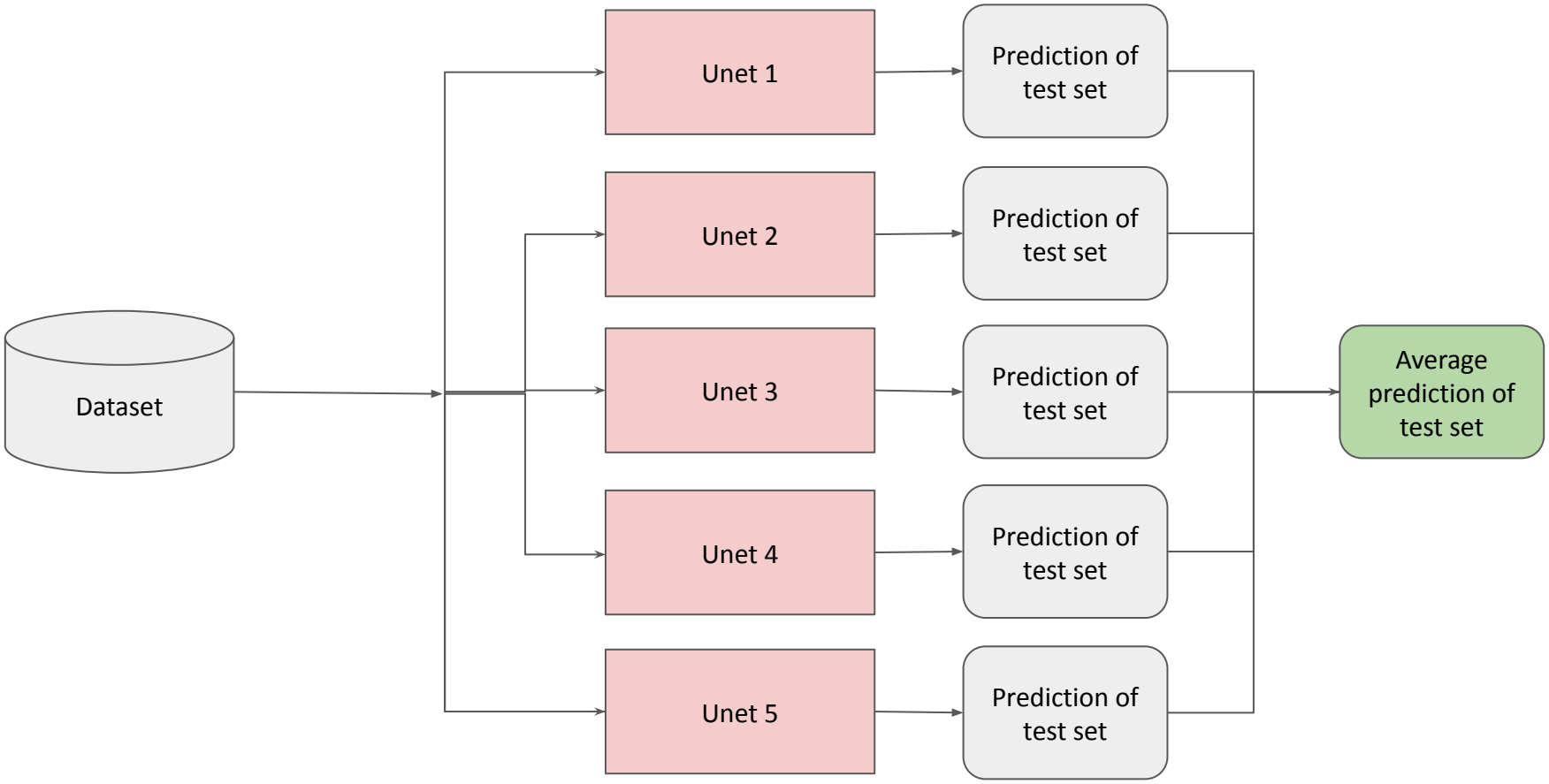
Summary: The DL-based approach for EPID dose reconstruction shows promise for clinical use, with high accuracy in some cases.

Key Challenges: Predicting dose distribution for **complex phantom** configurations remains a significant challenge.

Next Steps: Expand the **dataset**, improve the model's **accuracy** in difficult cases, and explore potential applications in clinical workflows.

lorenzo.marini@phd.unipi.it

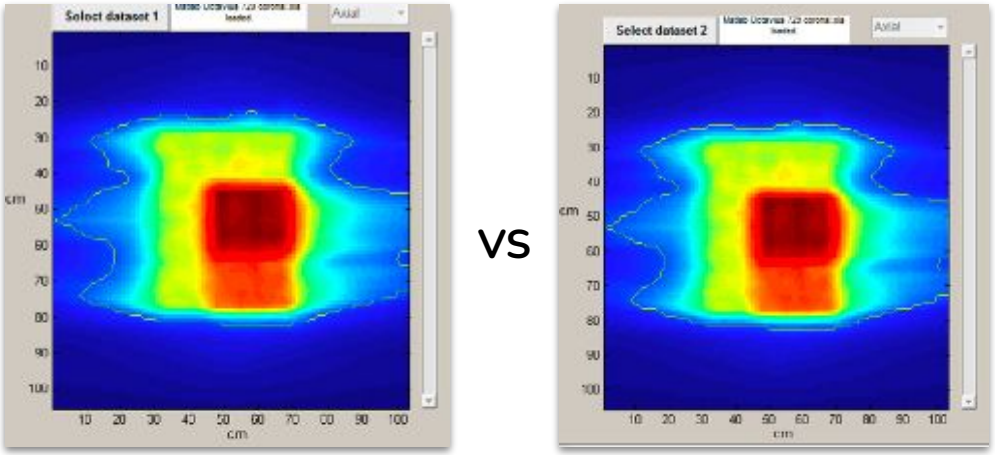
Cross validation & ensemble learning



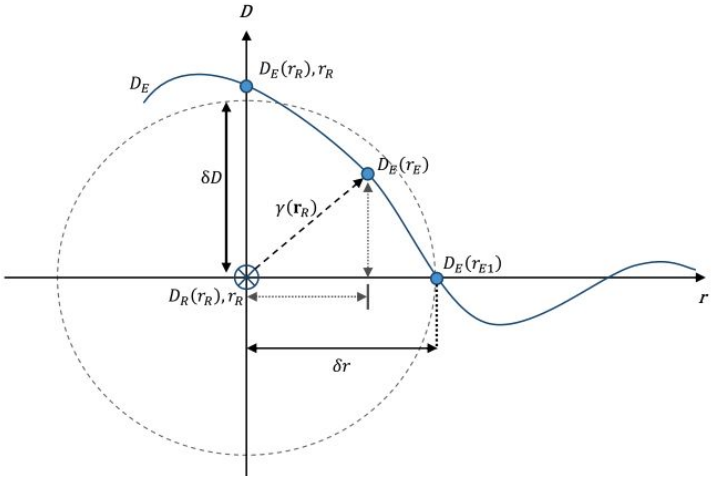
- CV is employed to ensure the model is **robust** and **generalizes well** to unseen data.
- Evaluate the performance on different subsets of the dataset, reducing the **risk of overfitting** and providing a more reliable estimate of its ability to predict new images.

How to evaluate the accuracy of the reconstruction?

Real **Reconstructed**



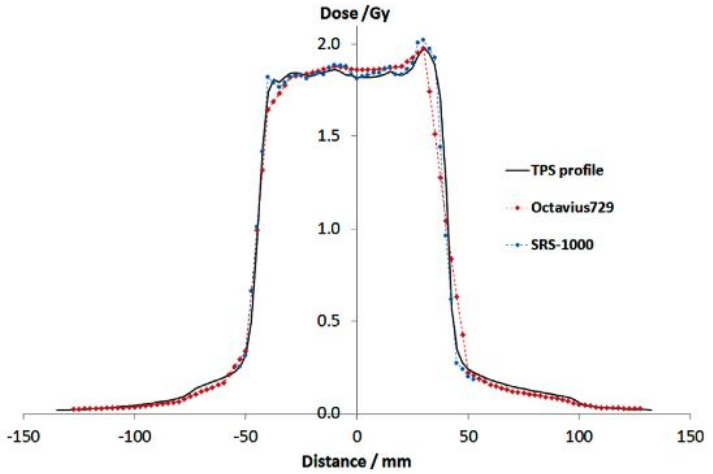
VS



$$\Gamma = \sqrt{\left(\frac{\text{Distance-to-Point}}{\text{Distance-Criterion}}\right)^2 + \left(\frac{\text{Dose-difference}}{\text{Dose-Criterion}}\right)^2}$$

$$\gamma(x, y) = min[\Gamma(x, y)]$$

- The agreement between simulated and reconstructed doses is assessed by **comparing both dose and position**.
- The **y-index** merges **dose difference** and **distance to agreement** (DTA) into one metric.
- The y-test creates a 2D space with dose difference (ΔD) and DTA (Δr) as axes.
- Points with **y > 1 fail the test**.
- A plane is **clinically acceptable** if over **90-95% of points have y ≤ 1**.



M. Hussein et al. / Physica Medica 36 (2017) 1–11

Limitations

Limited Data Availability → influence on the results' quality and reliability, limiting the developed model's generalizability.

Data Coming from a single Center → introducing a potential bias in the results.

Complexity of the problem → at the moment it is not possible to predict with certainty the level of accuracy that will be able to be achieved in the dose reconstruction on the 3D problem.

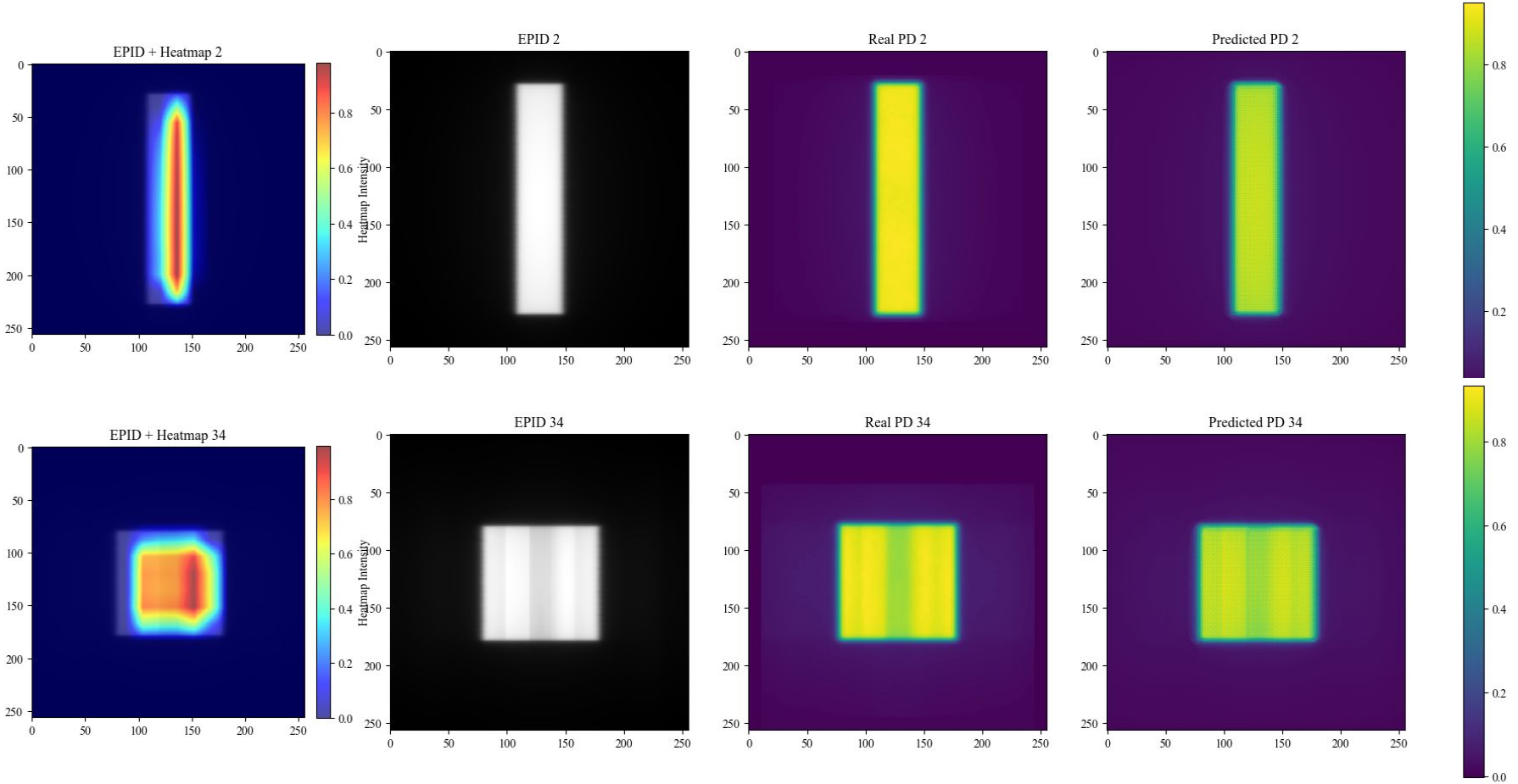
Evaluation of the Model Performance → Using metrics such as MAE or γ-index for model evaluation may not be the best choice.

Explainability → Developing an extremely complex algorithm could reduce its explainability. Finding a trade-off between performance and explainability will be required.

lorenzo.marini@phd.unipi.it

Explainability & interpretability

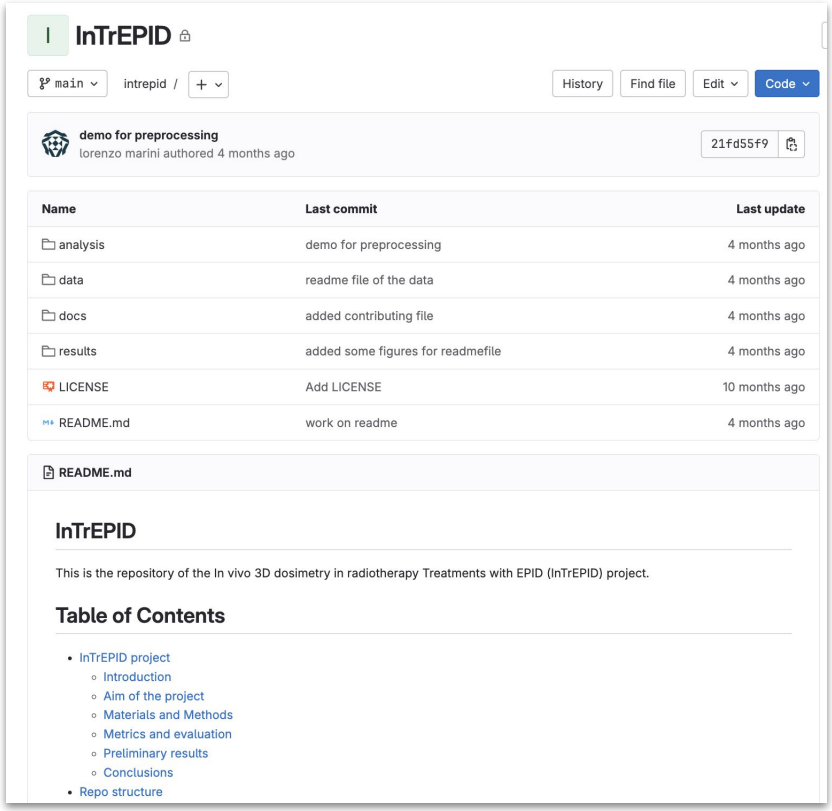
Visualizing what unet learn…

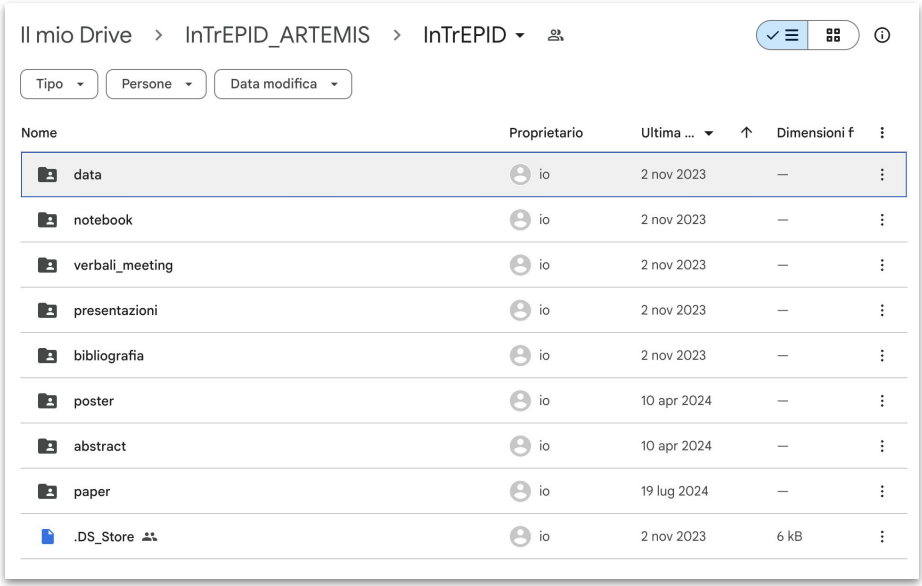


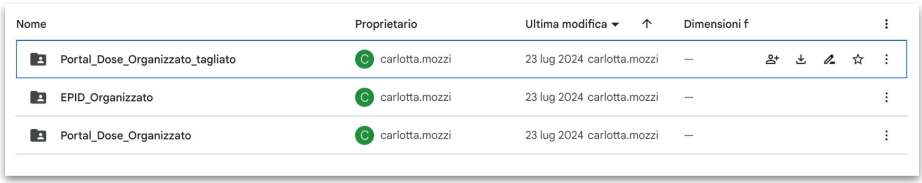
Heatmaps produced with Grad-CAM are a powerful tool for improving the **transparency** of the model, by helping to understand the model's decision pathway.

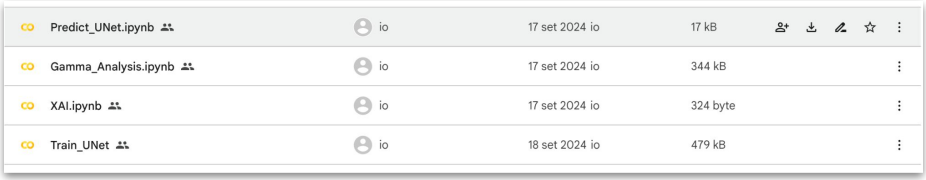
lorenzo.marini@phd.unipi.it

Sources: data & codes







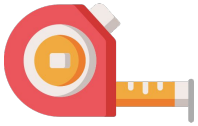


https://baltig.infn.it/lomarini/intrepid

InTrEPID_ARTEMIS

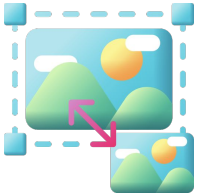
lorenzo.marini@phd.unipi.it

Data preprocessing



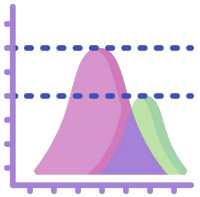
(1)
Matching of physical size and images centering

Do the images have the same dimensions and have the same physical position?



(2)
Image resizing (256x256)

A trade-off between resolution and computing capacity

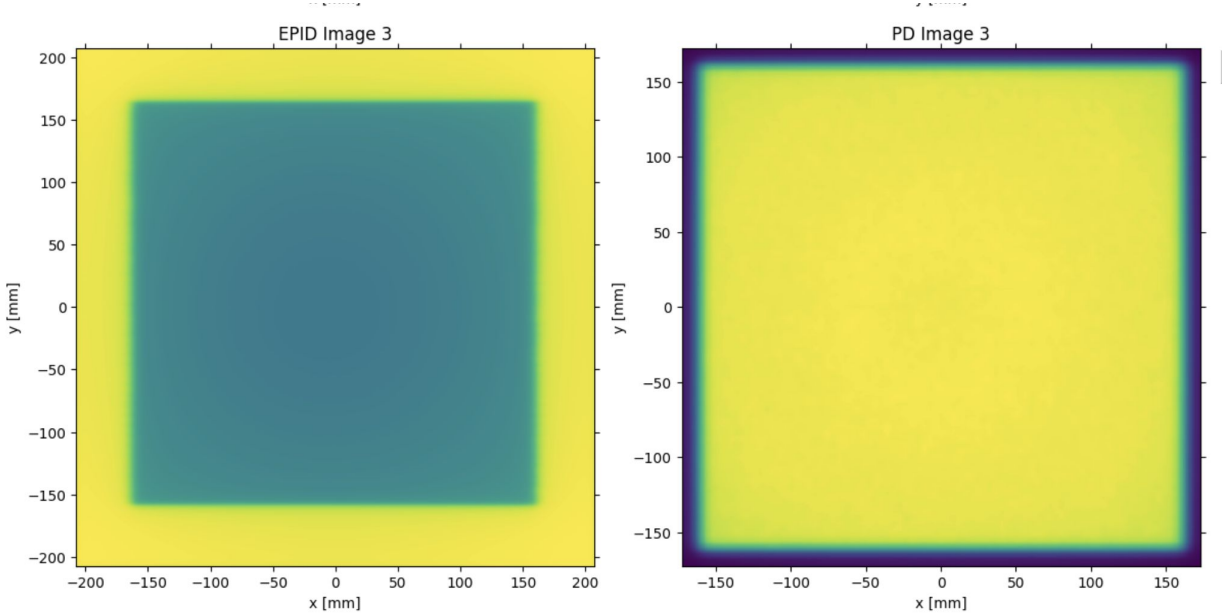


(3)
Pixel scaling → normalization

It can help the training phase of the NN

1) Matching of physical size

Do the images have the same dimensions?

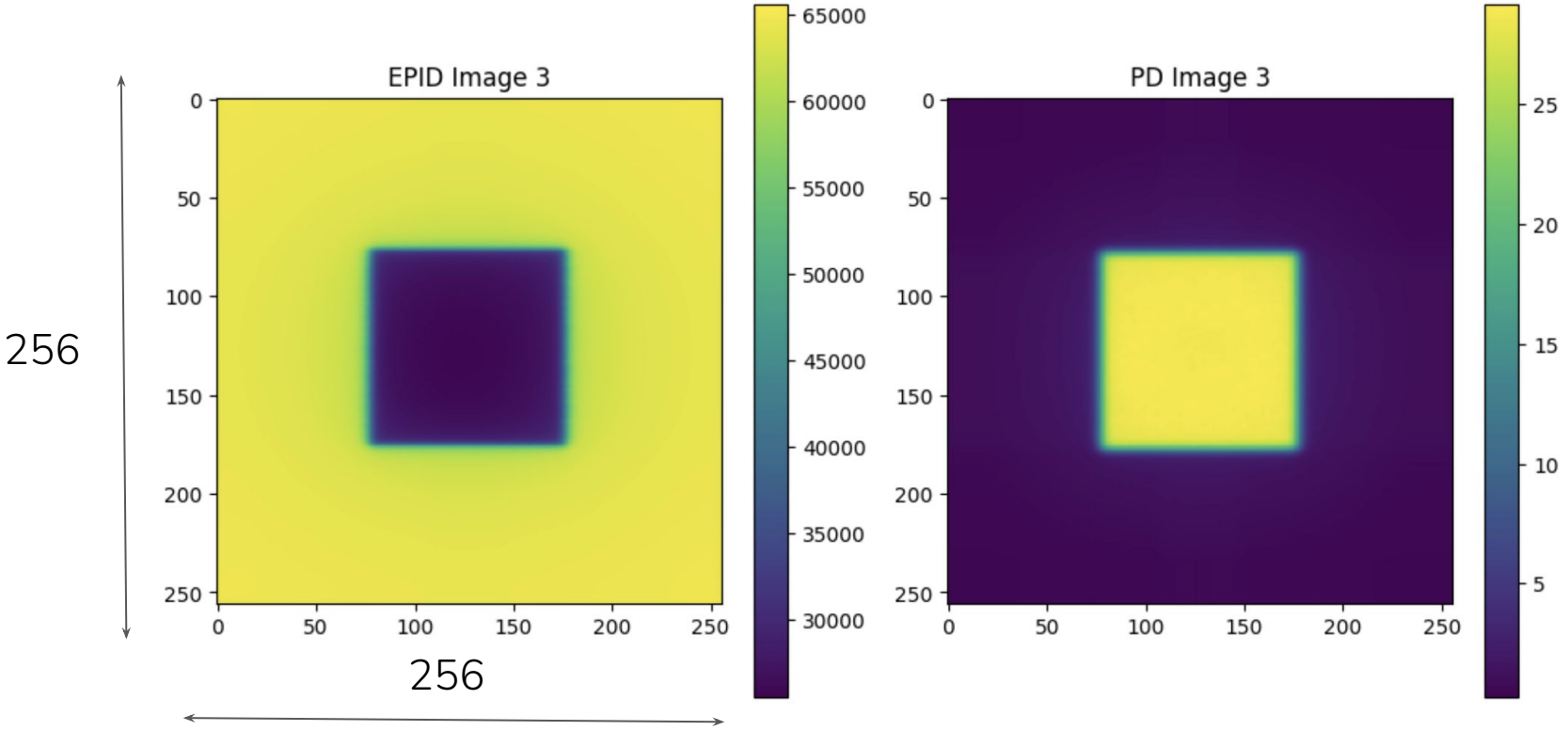


- Total dimensions are not the same
- Signals are not always perfectly centered

Is it better cut the EPID images or add background pixels to PDs?

2) Image resizing → (256x256)

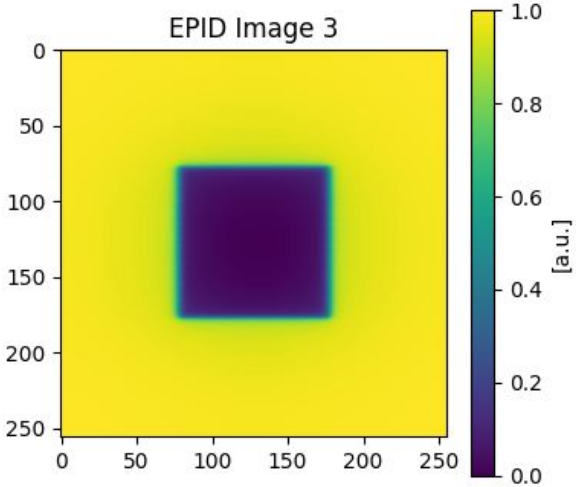
A trade-off between resolution and computing capacity

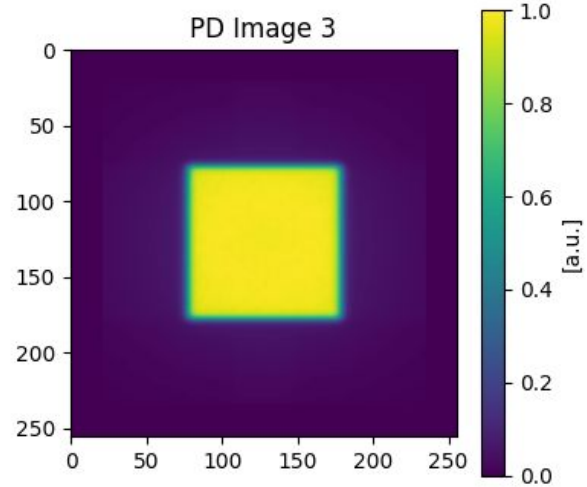


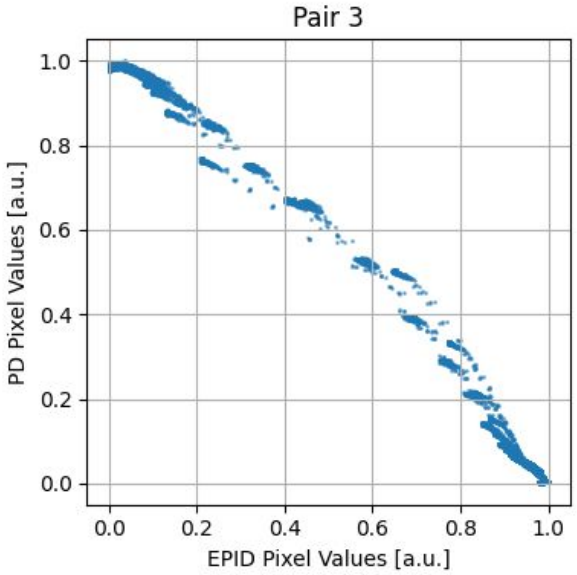
- **Image resizing** (using OpenCV, Python) refers to the scaling of images.
- Reduces the number of pixels, speeding up NN training and reducing model complexity (**Training efficiency**).
- Lowers computational and memory requirements by decreasing image size.

3) Pixel scaling

It can help the training phase of the network







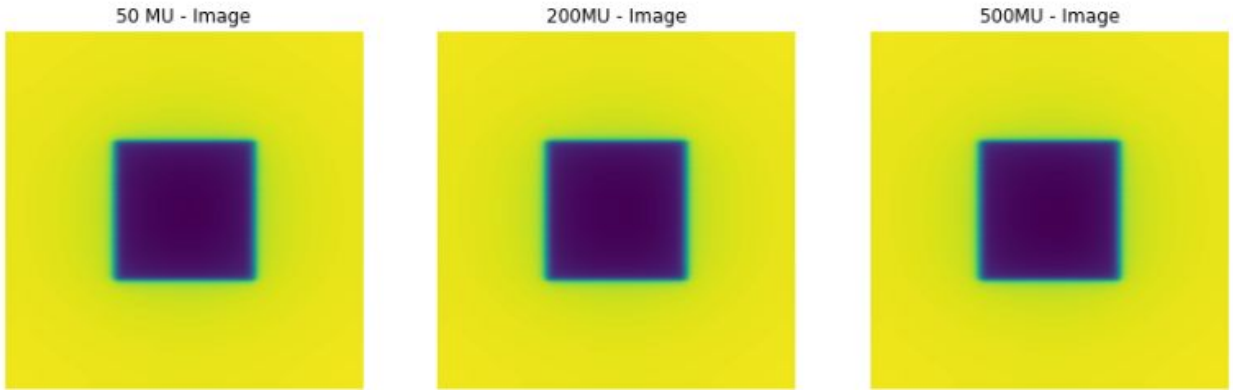
$EPID_i \rightarrow EPID_i/\max(EPID \text{ or } CV)$

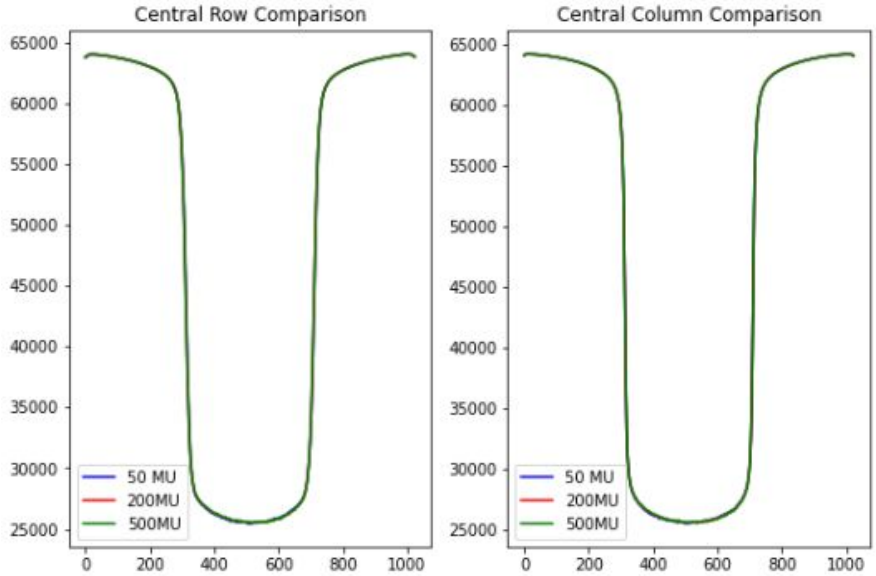
$PD_i \rightarrow PD_i/\max(PD \text{ or } CV)$

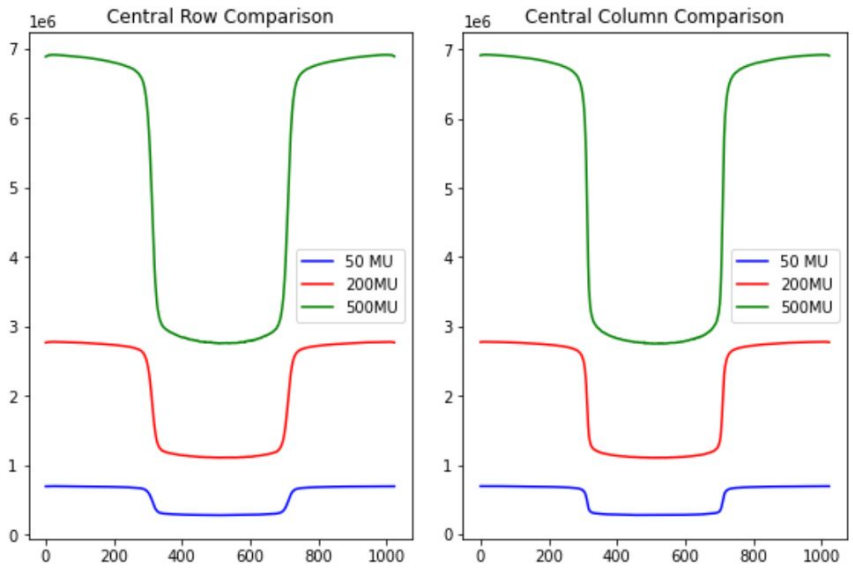
- Normalization is the process of converting an actual range of values which a numerical pixel can take, into a **standard range of values**, typically in the **interval [0, 1]**.
- ***Why do we normalize?*** It is not a strict requirement. However, in practice, it can lead to an **increased speed of learning** (Gradient descent, weight updates and numerical overflow)

lorenzo.marini@phd.unipi.it

EPID normalization factor



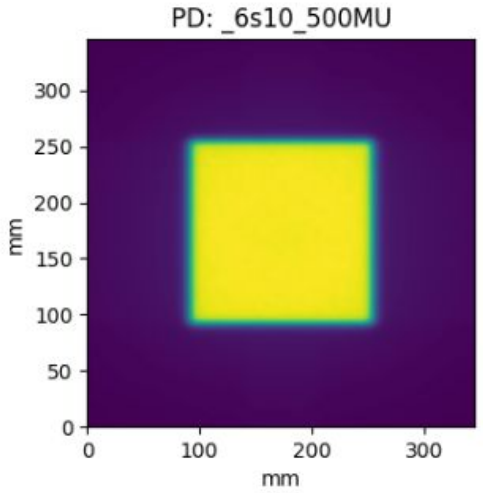


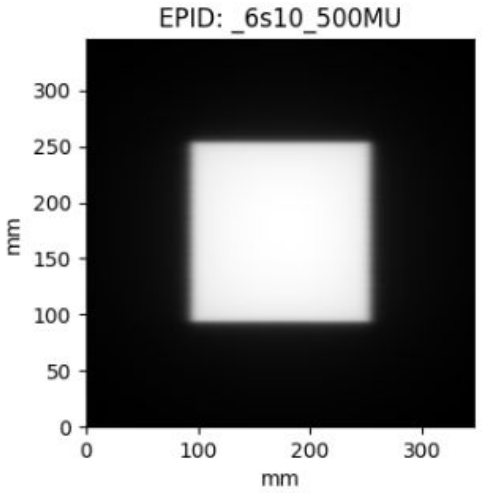


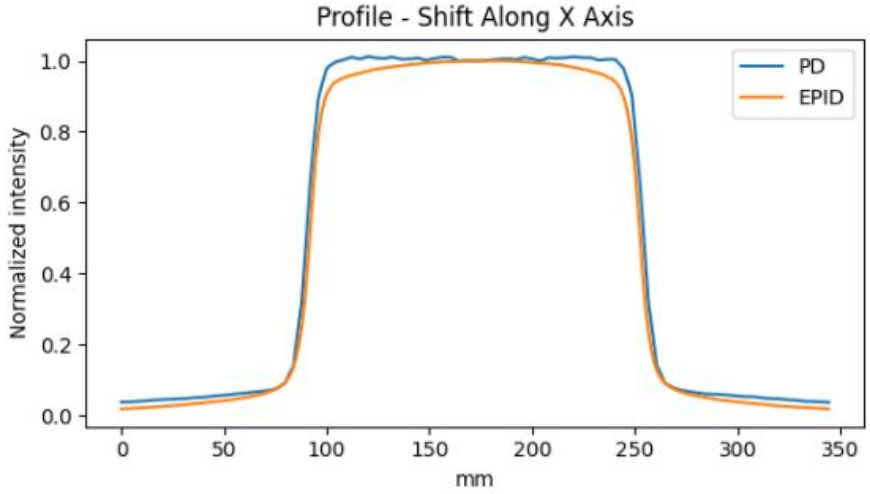
Divided by normalization factor
(tag 21x1002 DCOM file)

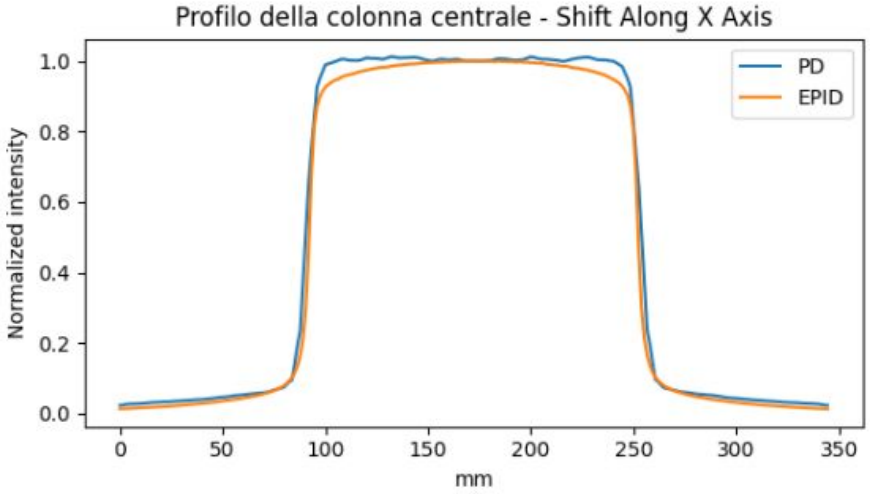
lorenzo.marini@phd.unipi.it

Preprocessing Results

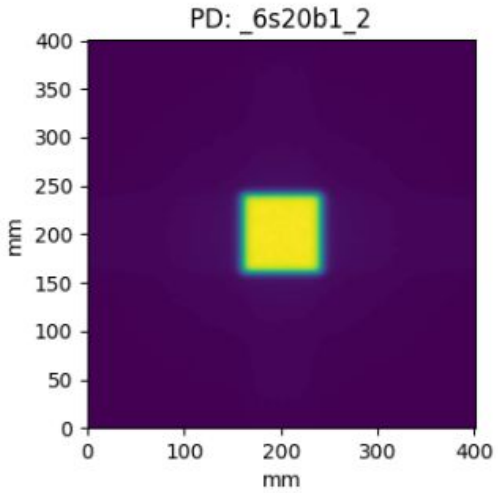


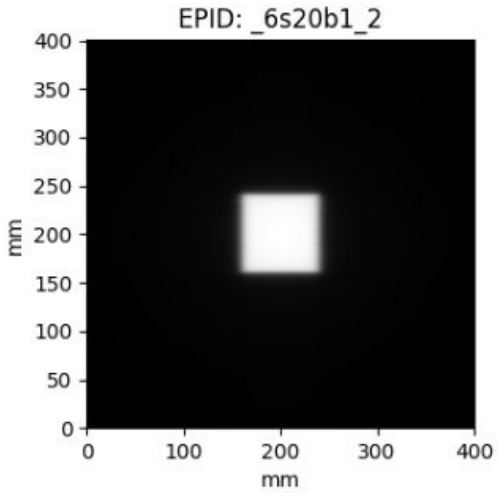


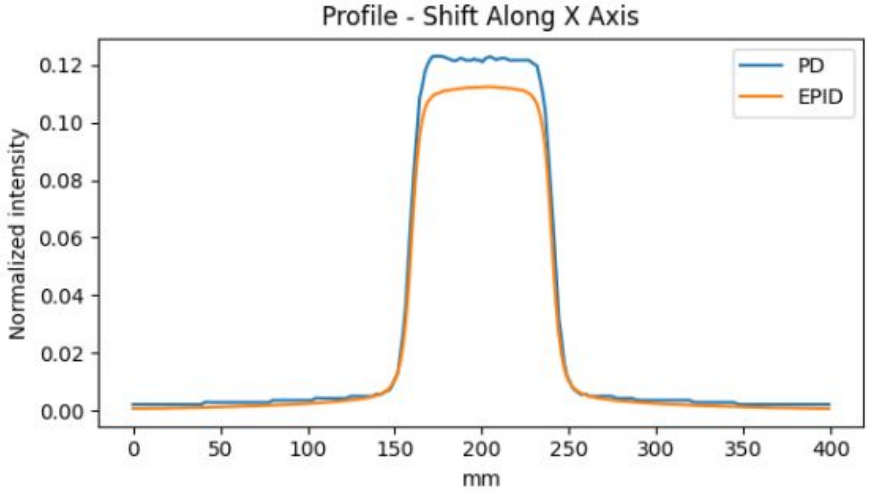


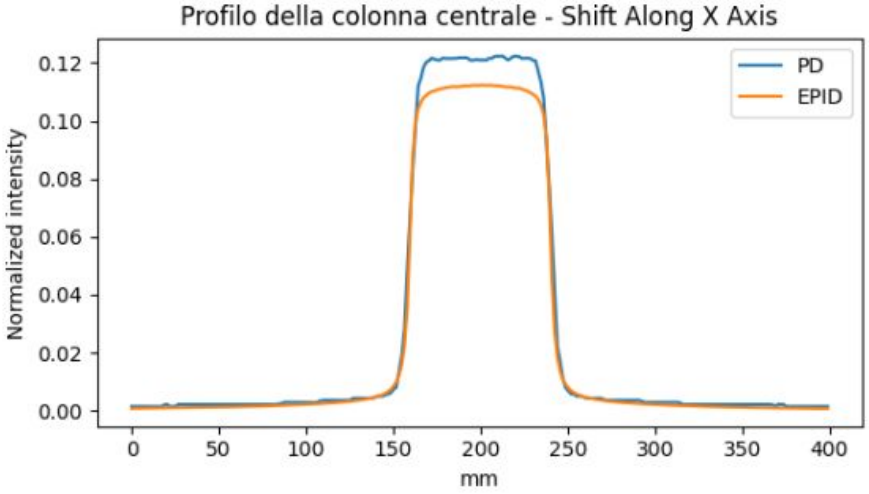


Preprocessing Results

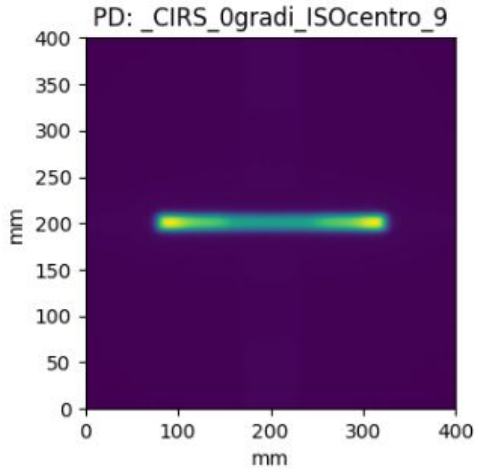


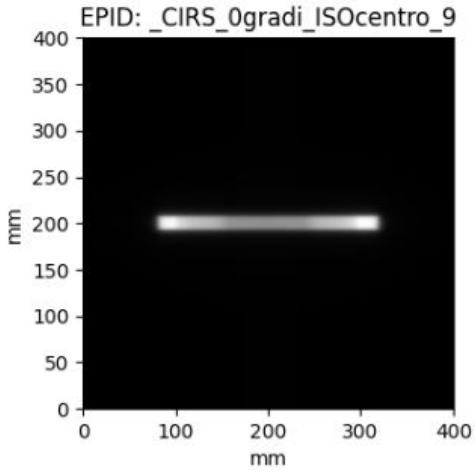


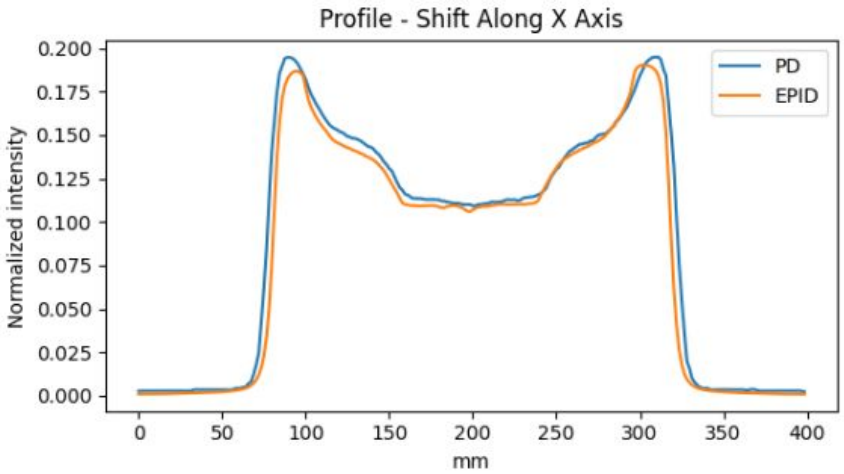


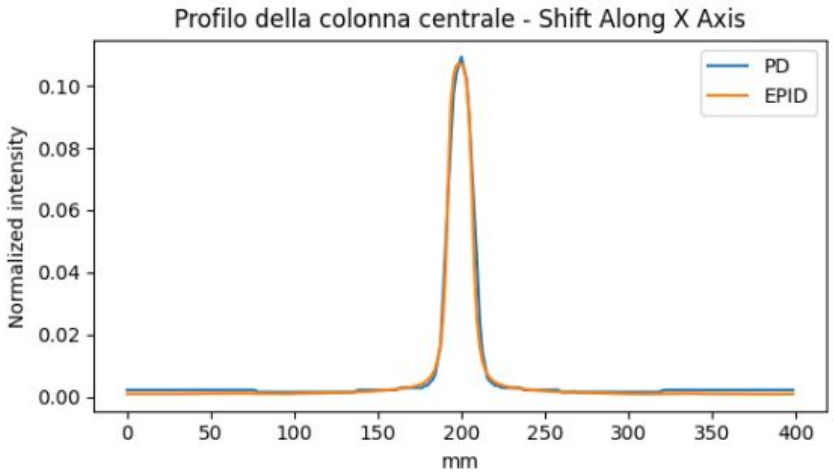


Preprocessing Results

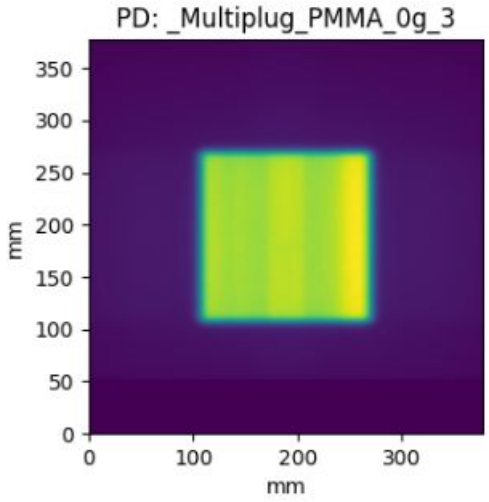


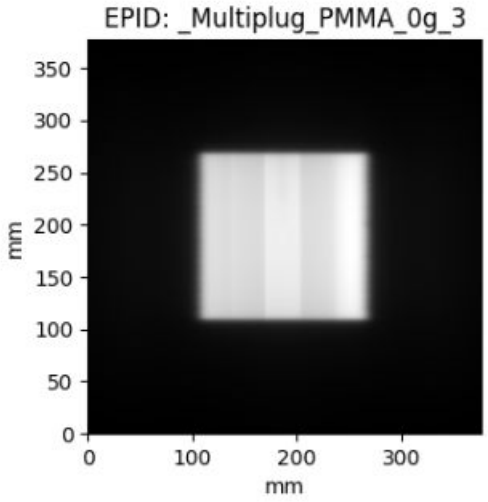


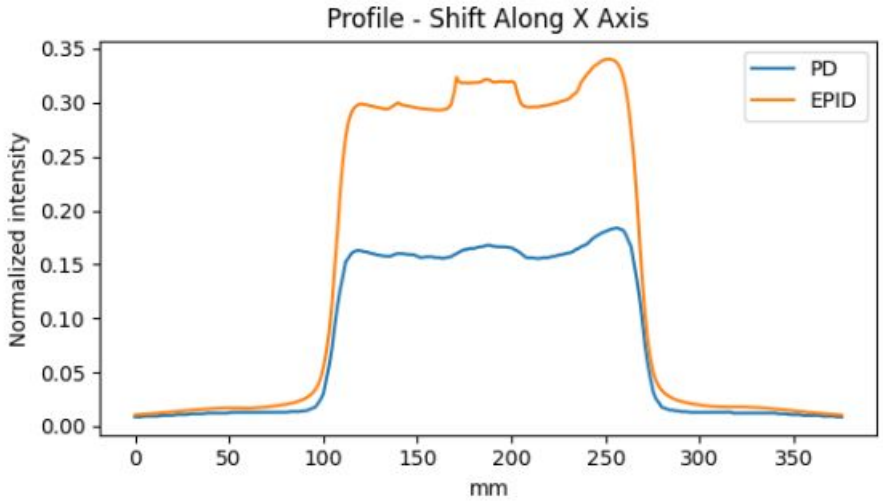


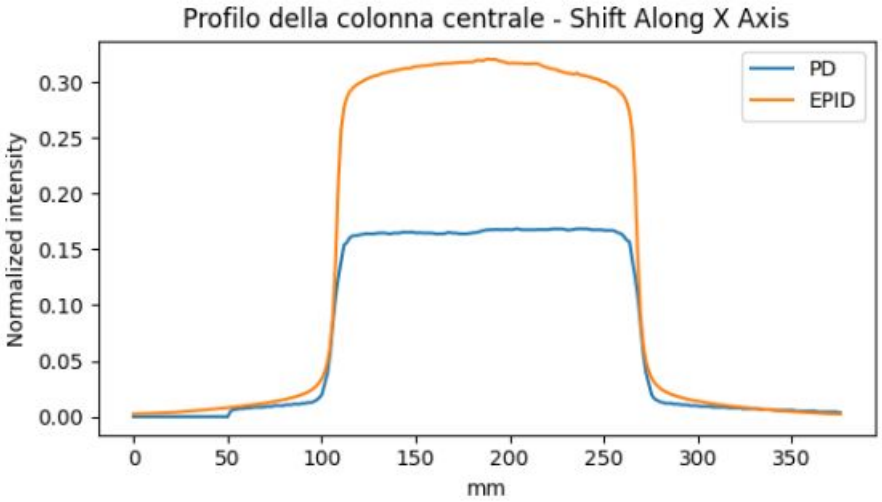


Preprocessing Results









lorenzo.marini@phd.unipi.it