KU LEUVEN

UZ LEUVEN

Uncertainty Quantification for Synthetic Medical Images

Paper 13928-24

SPIE Medical Imaging 2026

Konstantinos Koukoutegos[1,2], Elena Sizikova[3], Hilde Bosmans[1,2], Aldo Badano[3]

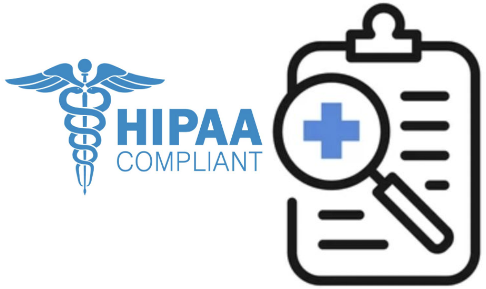
[1] UZ Leuven, Department of Radiology, Leuven, Belgium
[2] KU Leuven, Faculty of Medicine, Department of Imaging and Pathology, Division of Medical Physics & Quality Assessment, Leuven, Belgium
[3] U.S. Food and Drug Administration, Center for Devices and Radiological Health, Office of Science and Engineering Laboratories, Silver Spring, Maryland, USA

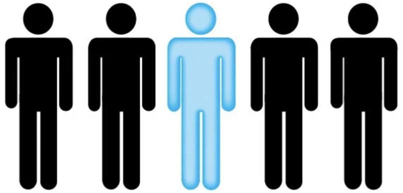
Disclaimer

This project reflects the views of the authors and does not represent the views or policy of the U.S. Food and Drug Administration, the Department of Health and Human Services, or the U.S. Government. The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.

Background and Motivation

Synthetic medical data





disease coverage

annotation burden

Standard evaluations focus on mean performance
Two models with similar evaluation metrics may have very different reliability
No established paradigm for uncertainty induced specifically by synthetic data

Background and Motivation

Types of uncertainty in AI

Aleatoric	***Epistemic***
Inherent randomness or noise in a system	Lack of knowledge about the model parameters or the data generating process
Irreducible	Can be explained away with better modelling or more data

Background and Motivation

Types of uncertainty in AI

Aleatoric	***Epistemic***
Inherent randomness or noise in a system	Lack of knowledge about the model parameters or the data generating process
Irreducible	Can be explained away with better modelling or more data

L. Huang, S. Ruan, Y. Xing, and M. Feng. *A review of uncertainty quantification in medical image analysis: Probabilistic and non-probabilistic methods.* Medical Image Analysis, 97:103223, 2024

Types of Uncertainty

Data uncertainty - emerging when models are trained on data from different source domains (e.g. patient and synthetic).

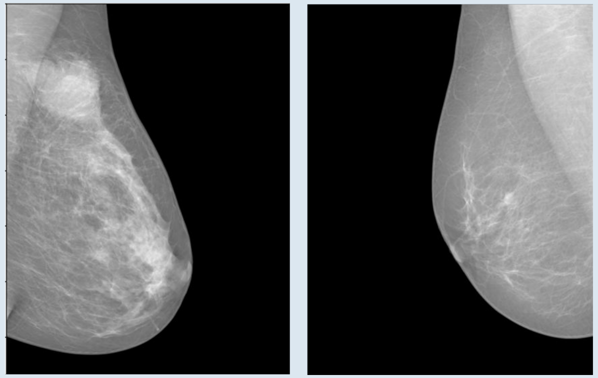
Reader uncertainty - which quantifies the variability across models trained on the same data but with different random initializations (e.g., random seeds). This form of epistemic uncertainty captures sensitivity to the learning process itself (also known as model or observer uncertainty).

Case uncertainty - captures variability in predictions due to differences across individual test cases, which reflects epistemic uncertainty at the input level (or subject uncertainty).

Methodology

Datasets

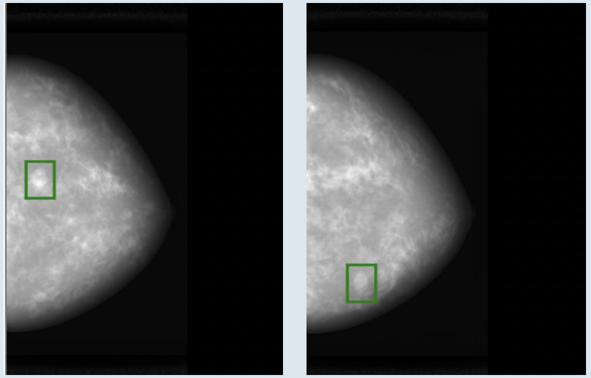
INbreast



publicly available
105 mammograms with lesions
ground truth lesion annotations

$$\mathcal{D}_{patient} = (\mathbf{X}_{patient}, \mathbf{Y}_{patient})$$

M-SYNTH



publicly available
1200 stochastic knowledged-based models
generated using the VICTRE pipeline
fatty breasts
lesion density of 1.0 and size 5mm
100% of clinically recommended dose

$$\mathcal{D}_{synthetic} = (\mathbf{X}_{synthetic}, \mathbf{Y}_{synthetic})$$

Methodology

$$\mathcal{D}_{p,q} = (p\mathcal{D}_{patient}) \cup (q\mathcal{D}_{synthetic})$$

$$\mathcal{D}_{p,q} = (p\mathbf{X}_{patient}, p\mathbf{Y}_{patient}) \cup (q\mathbf{X}_{synthetic}, q\mathbf{Y}_{synthetic})$$

$$\mathcal{D}_{p,q} = ((p\mathbf{X}_{patient} \cup q\mathbf{X}_{synthetic}), (p\mathbf{Y}_{patient} \cup q\mathbf{Y}_{synthetic}))$$

Methodology

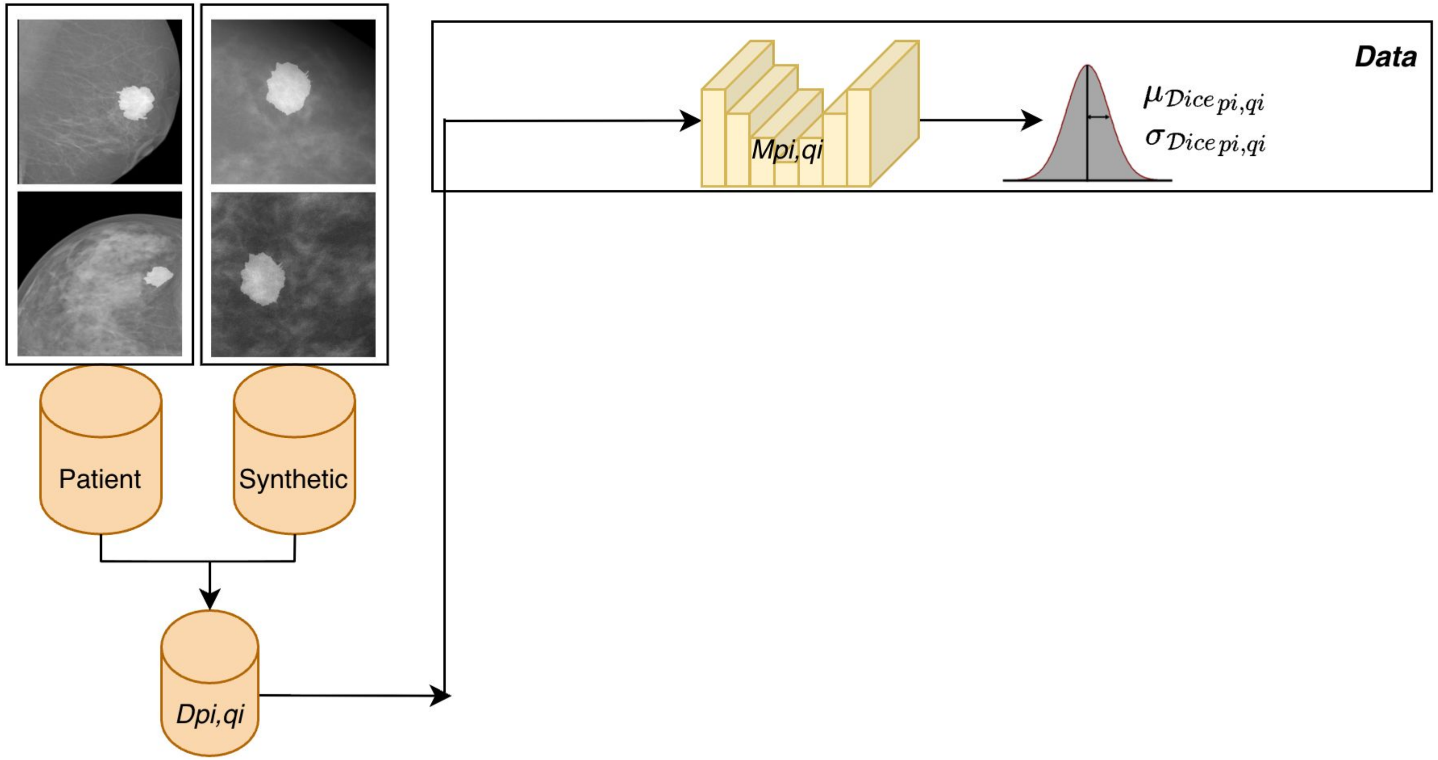
$$\mathcal{D}_{p,q} = (p\mathcal{D}_{patient}) \cup (q\mathcal{D}_{synthetic})$$

$$\mathcal{D}_{p,q} = (p\mathbf{X}_{patient}, p\mathbf{Y}_{patient}) \cup (q\mathbf{X}_{synthetic}, q\mathbf{Y}_{synthetic})$$

$$\mathcal{D}_{p,q} = ((p\mathbf{X}_{patient} \cup q\mathbf{X}_{synthetic}), (p\mathbf{Y}_{patient} \cup q\mathbf{Y}_{synthetic}))$$

Data

Standard neural net



Methodology

$$\mathcal{D}_{p,q} = (p\mathcal{D}_{patient}) \cup (q\mathcal{D}_{synthetic})$$

$$\mathcal{D}_{p,q} = (p\mathbf{X}_{patient}, p\mathbf{Y}_{patient}) \cup (q\mathbf{X}_{synthetic}, q\mathbf{Y}_{synthetic})$$

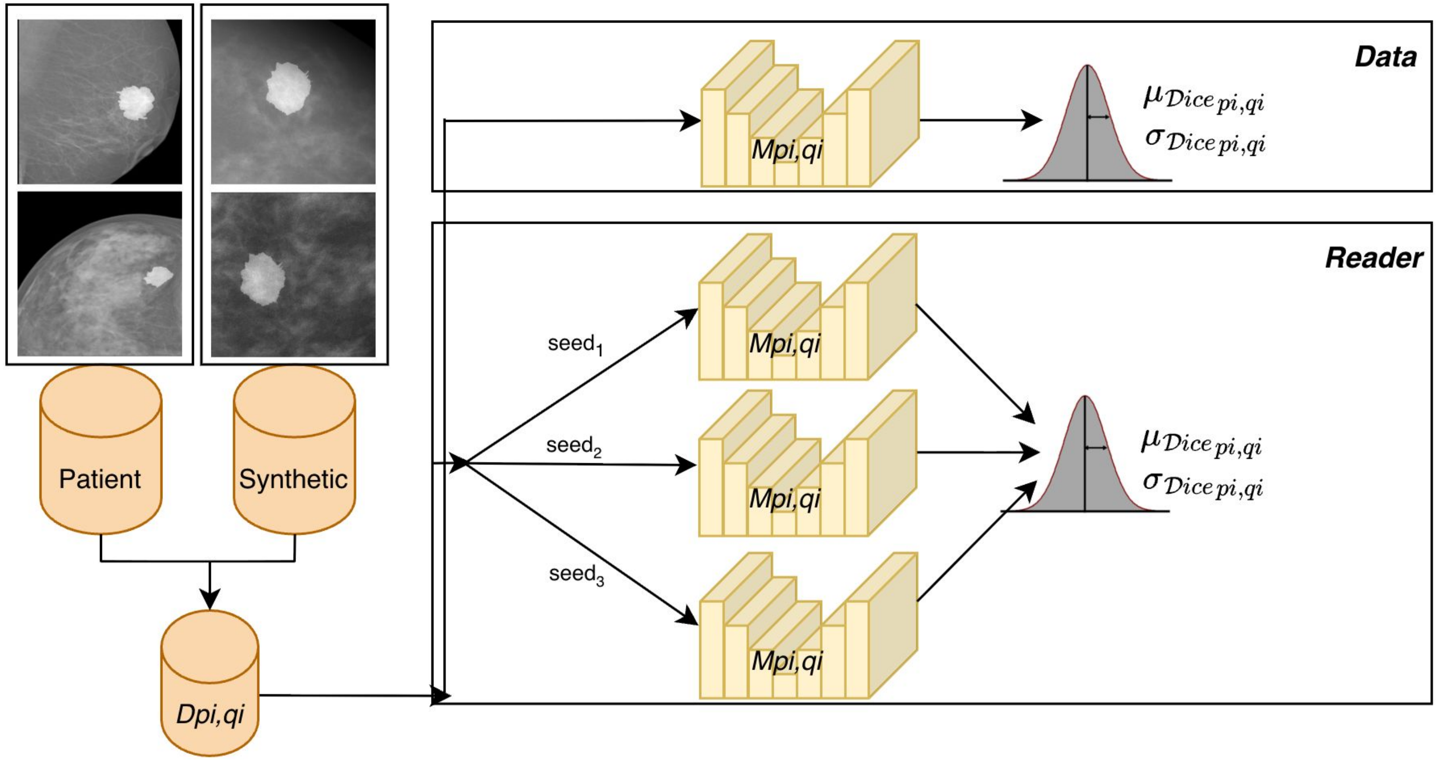
$$\mathcal{D}_{p,q} = ((p\mathbf{X}_{patient} \cup q\mathbf{X}_{synthetic}), (p\mathbf{Y}_{patient} \cup q\mathbf{Y}_{synthetic}))$$

Data

Standard neural net

Reader

Model ensembling



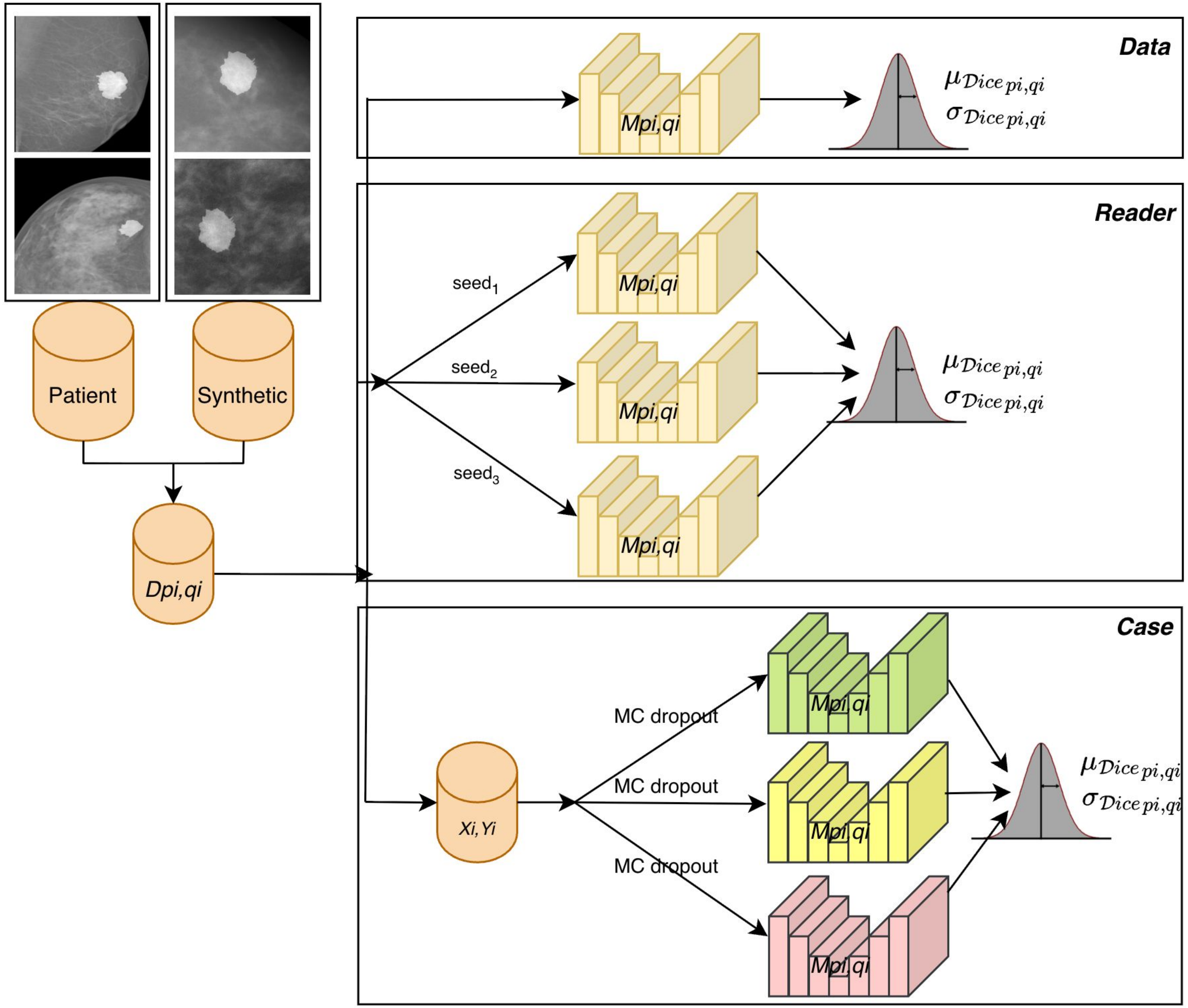
Methodology

$$\mathcal{D}_{p,q} = (p\mathcal{D}_{patient}) \cup (q\mathcal{D}_{synthetic})$$

$$\mathcal{D}_{p,q} = (p\mathbf{X}_{patient}, p\mathbf{Y}_{patient}) \cup (q\mathbf{X}_{synthetic}, q\mathbf{Y}_{synthetic})$$

$$\mathcal{D}_{p,q} = ((p\mathbf{X}_{patient} \cup q\mathbf{X}_{synthetic}), (p\mathbf{Y}_{patient} \cup q\mathbf{Y}_{synthetic}))$$

Data	***Reader***	***Case***
Standard neural net	Model ensembling	Monte Carlo dropout

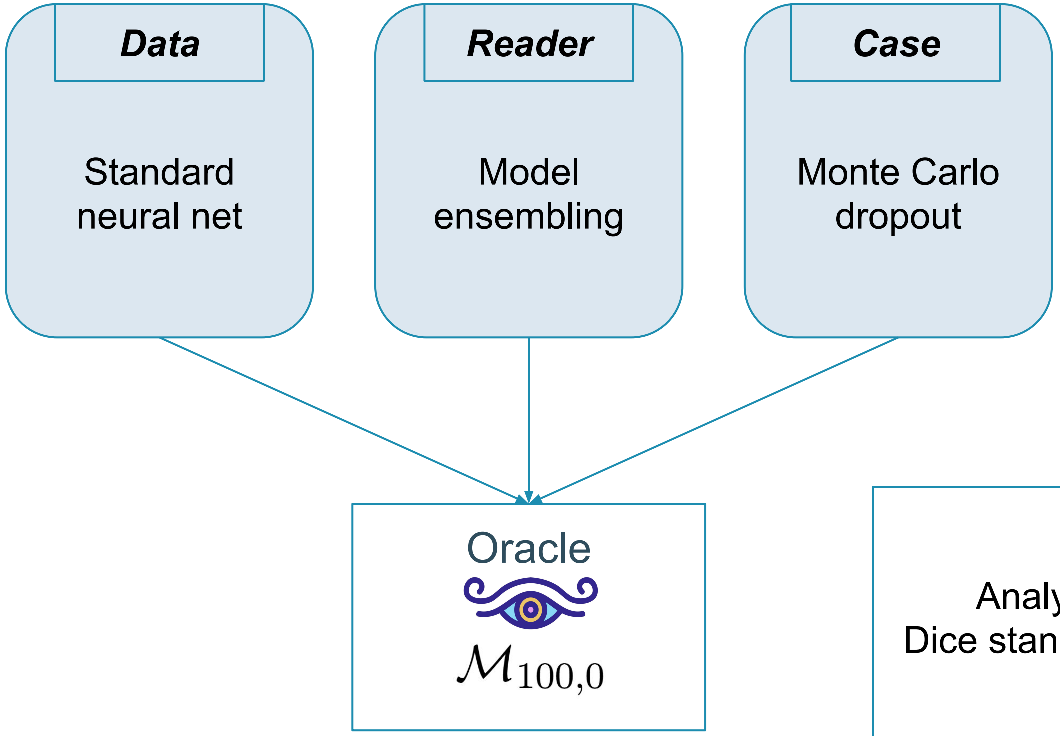


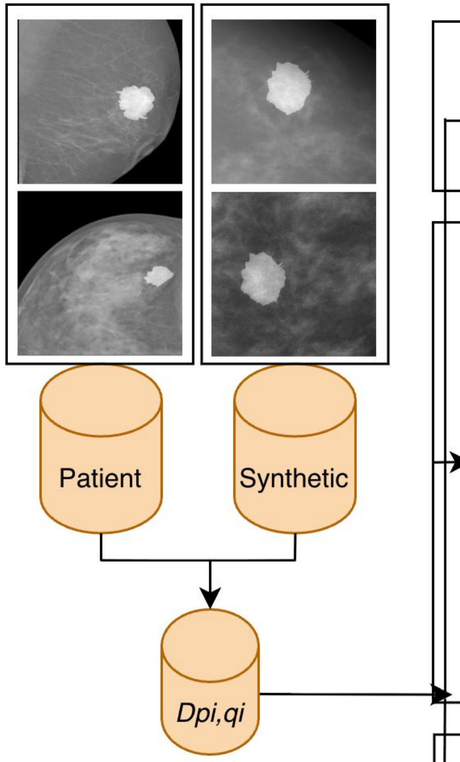
Methodology

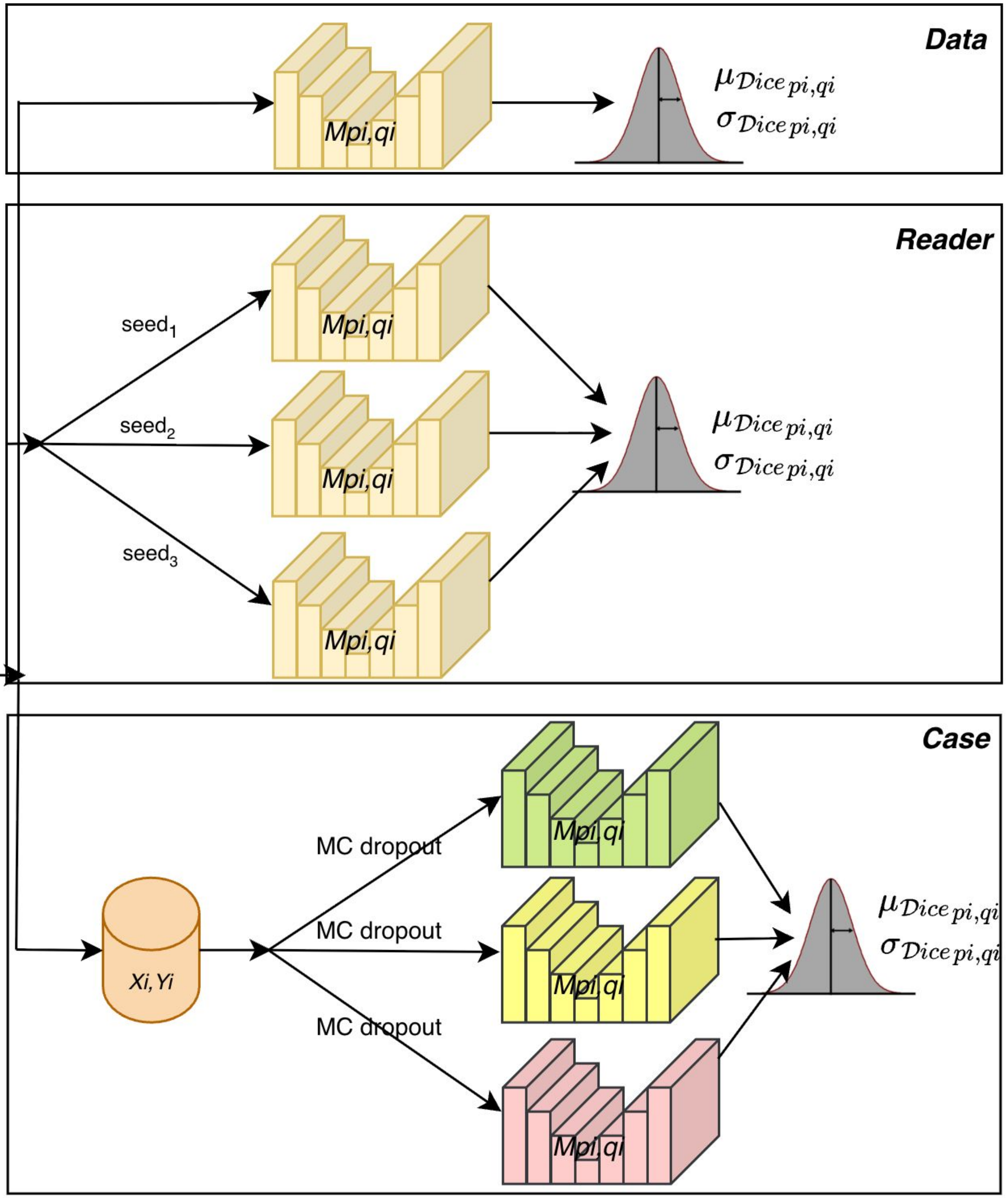
$$\mathcal{D}_{p,q} = (p\mathcal{D}_{patient}) \cup (q\mathcal{D}_{synthetic})$$

$$\mathcal{D}_{p,q} = (p\mathbf{X}_{patient}, p\mathbf{Y}_{patient}) \cup (q\mathbf{X}_{synthetic}, q\mathbf{Y}_{synthetic})$$

$$\mathcal{D}_{p,q} = ((p\mathbf{X}_{patient} \cup q\mathbf{X}_{synthetic}), (p\mathbf{Y}_{patient} \cup q\mathbf{Y}_{synthetic}))$$



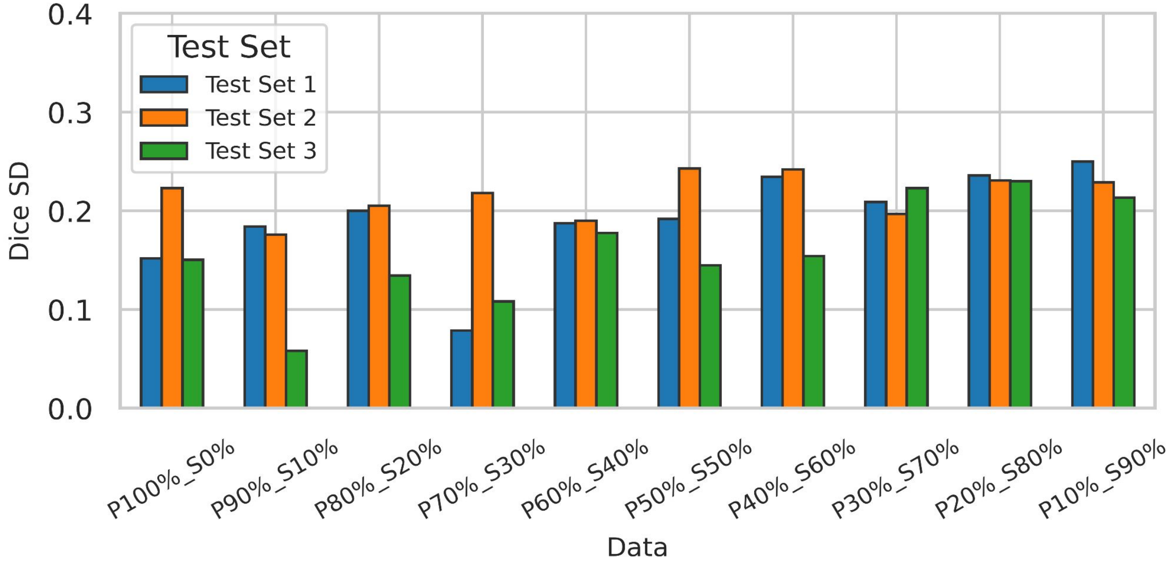




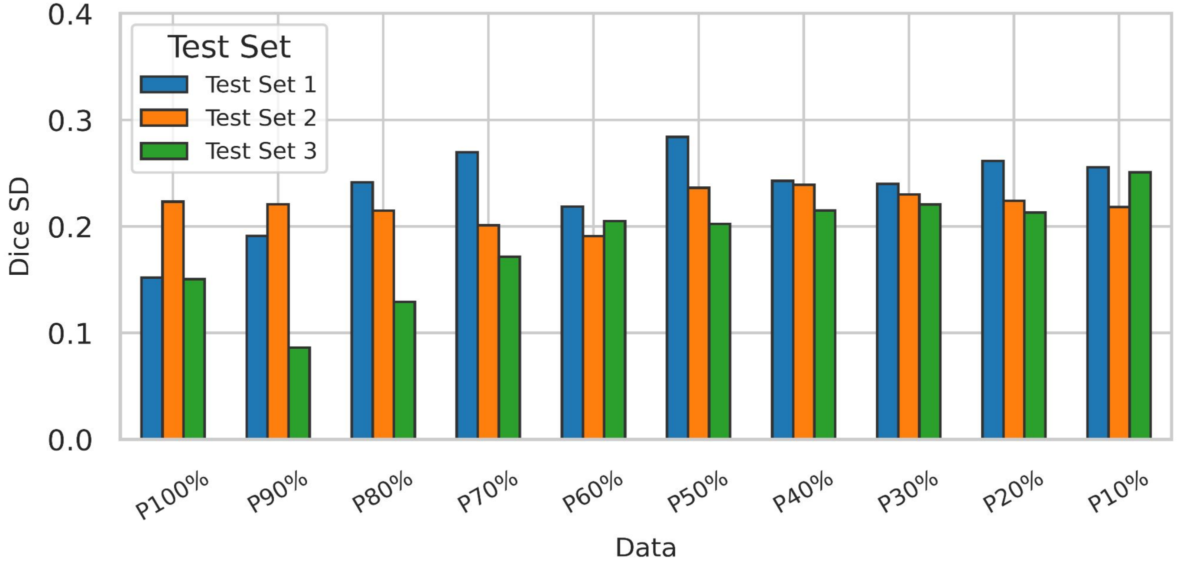
Analysis done using
Dice standard deviation (SD)

Results

Data Uncertainty



Patient:Synthetic (decreasing patient - increasing synthetic)



Patient:Synthetic (decreasing patient - no synthetic data)

UZ LEUVEN

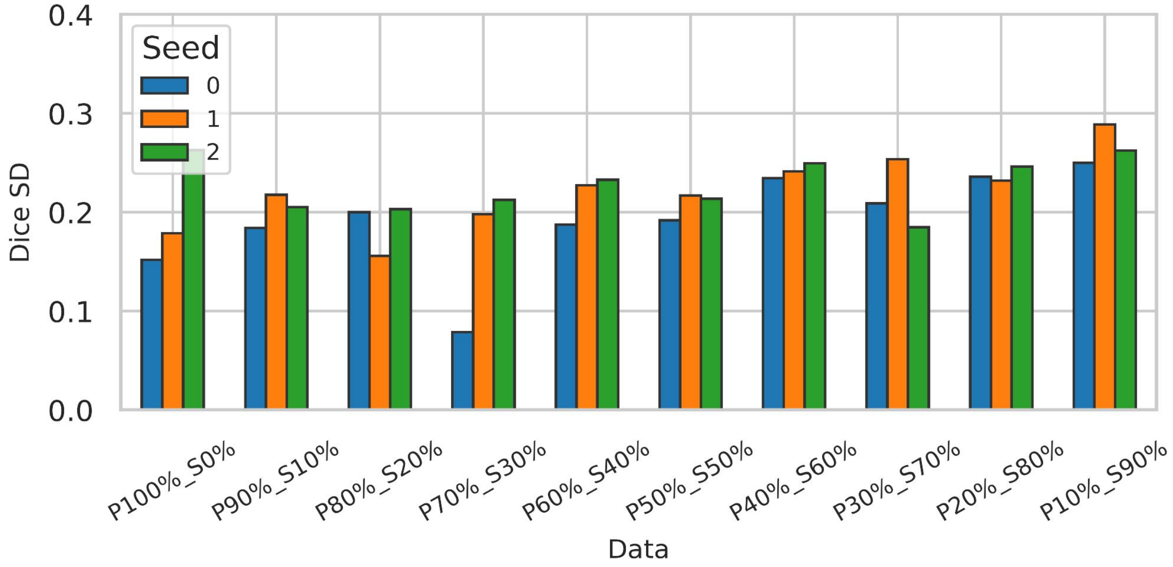
Department of Radiology

Faculty of Medicine, Department of Imaging and Pathology

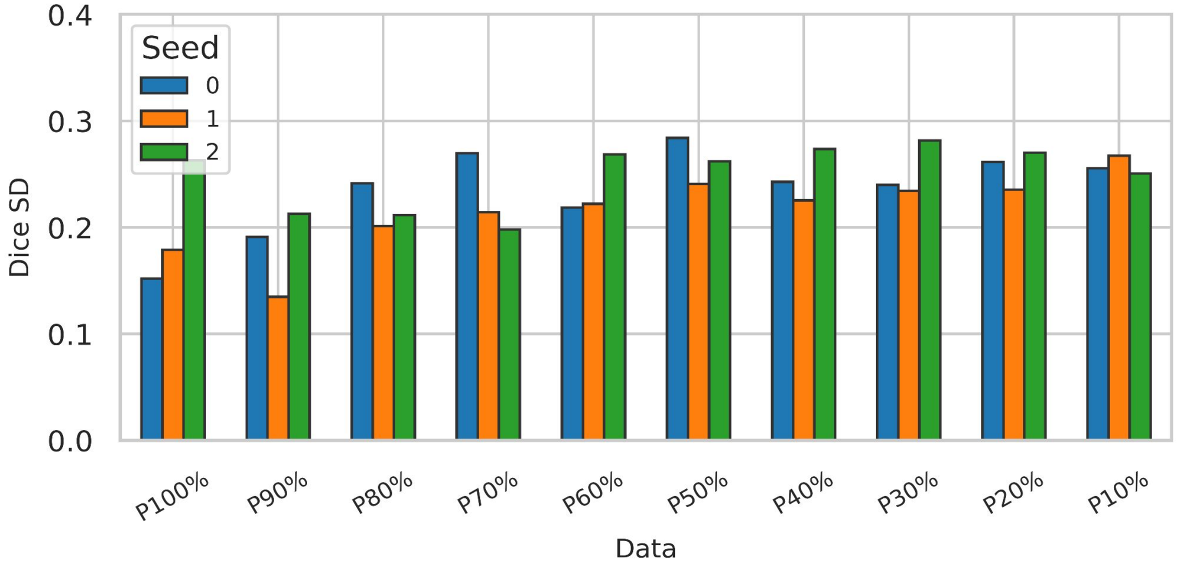
KU LEUVEN

Results

Reader Uncertainty



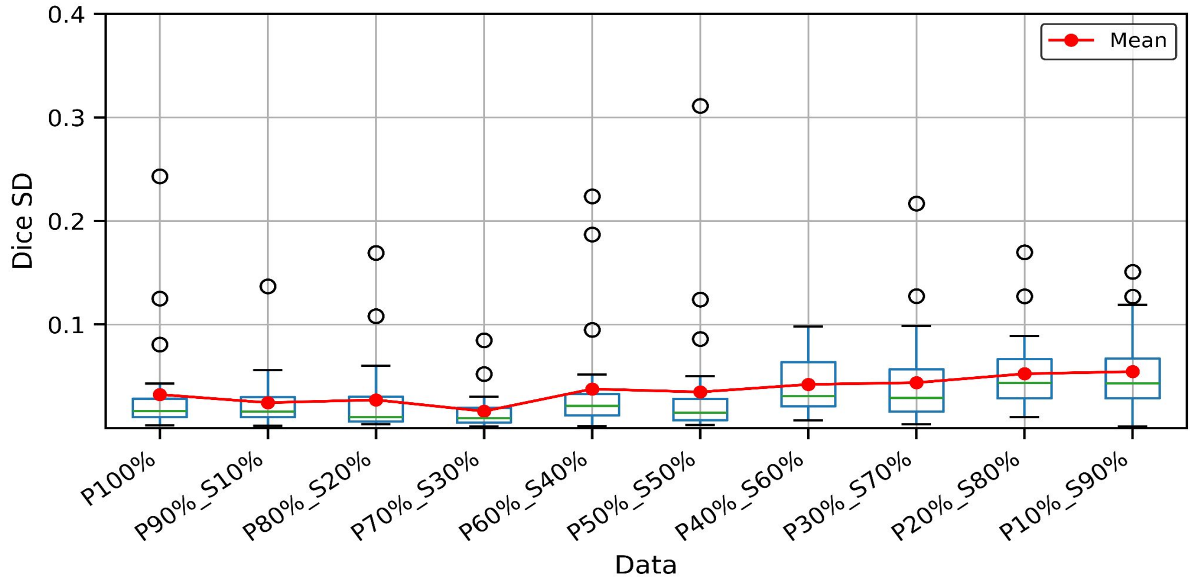
Patient:Synthetic (decreasing patient - increasing synthetic)



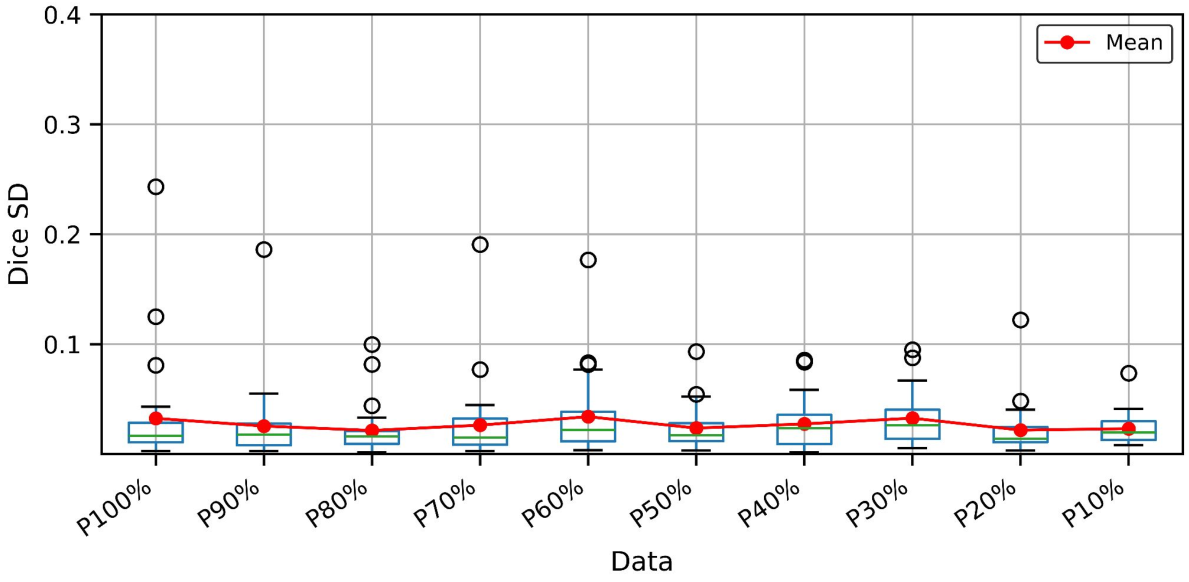
Patient:Synthetic (decreasing patient - no synthetic data)

Results

Case Uncertainty



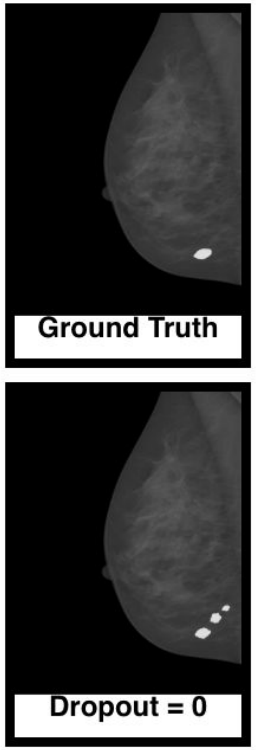
Patient:Synthetic (decreasing patient - increasing synthetic)

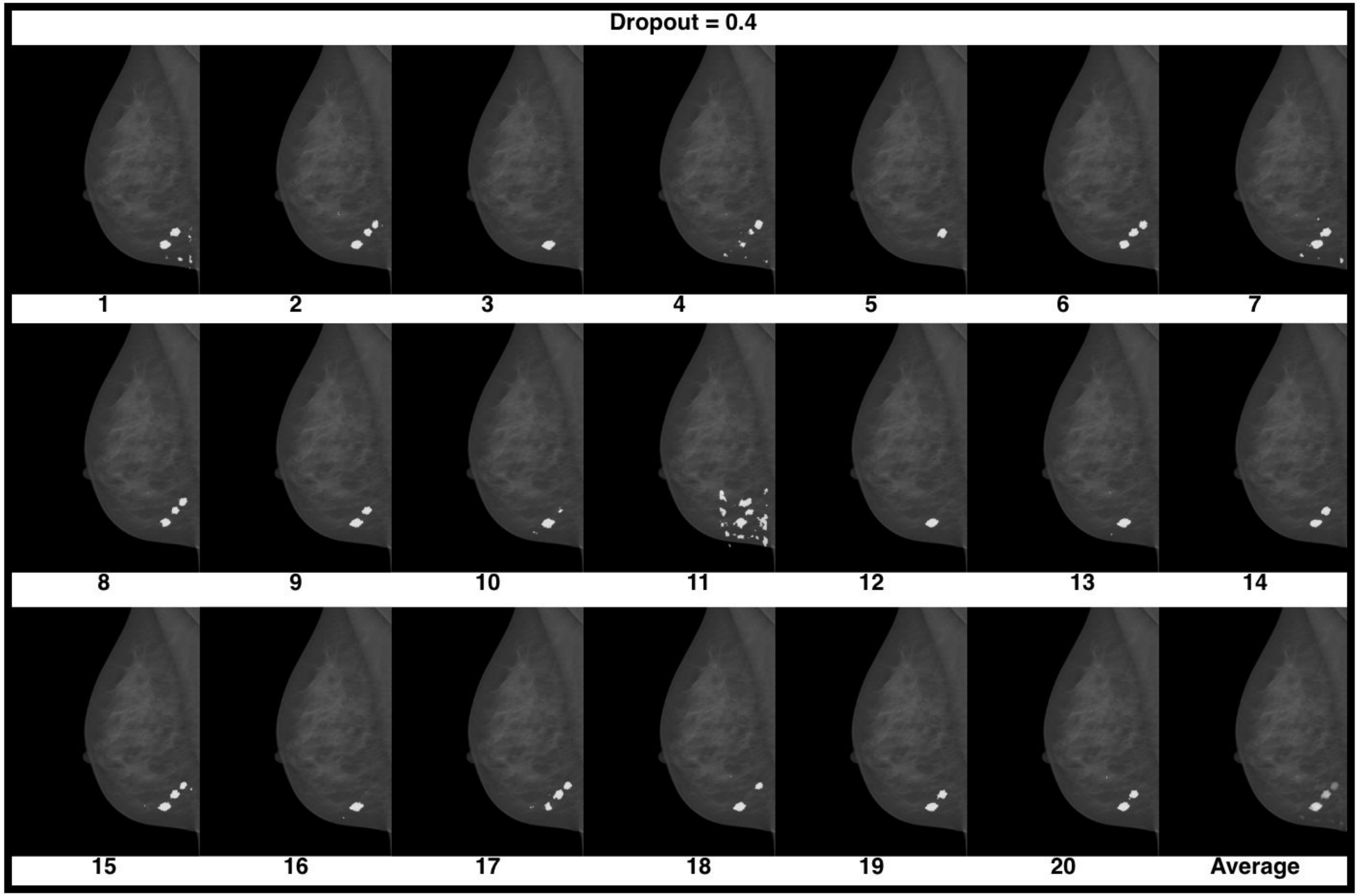


Patient:Synthetic (decreasing patient - no synthetic data)

Results

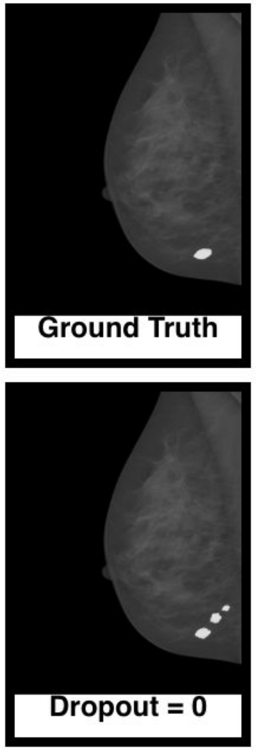
Case uncertainty

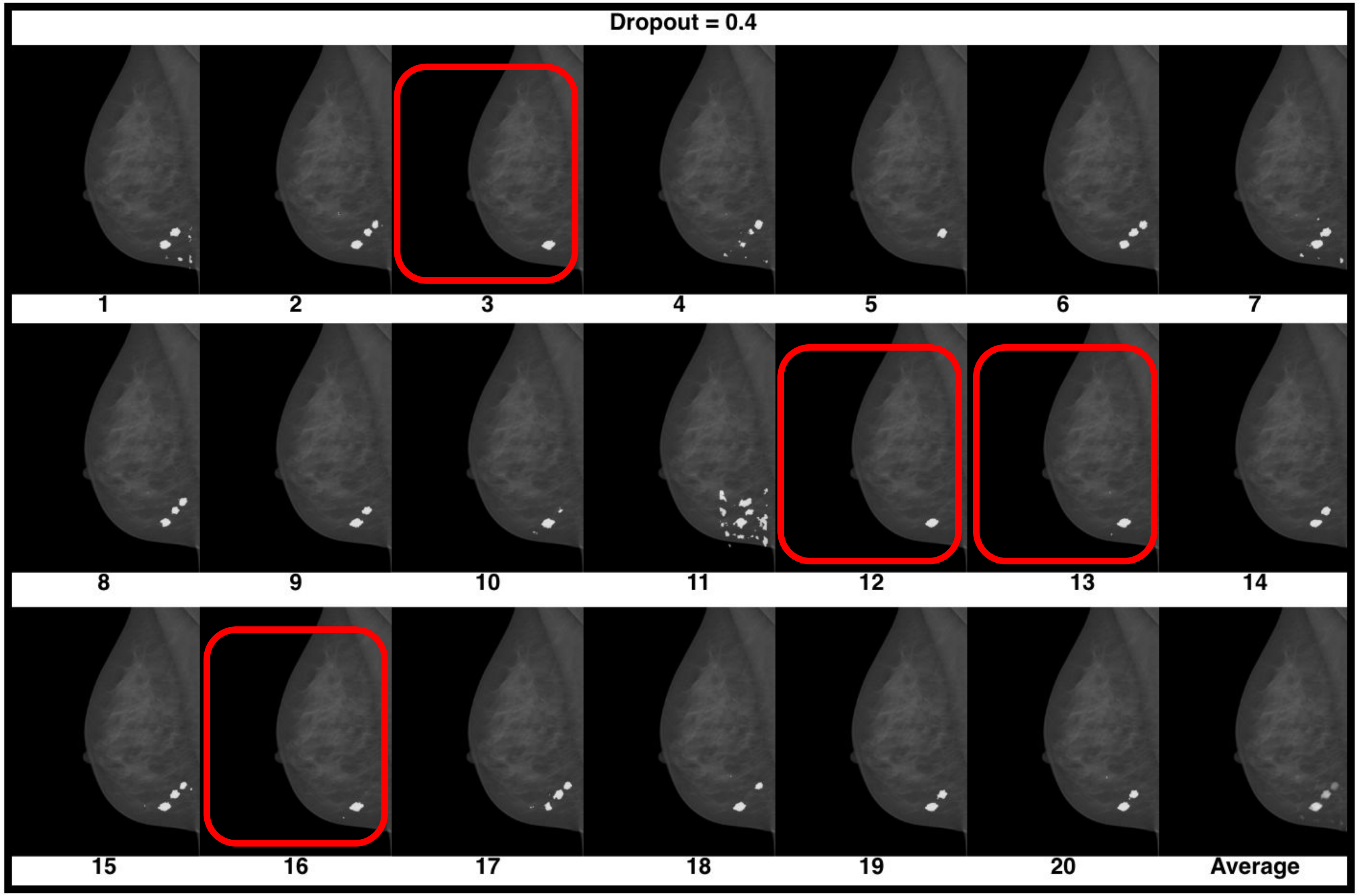




Results

Case uncertainty



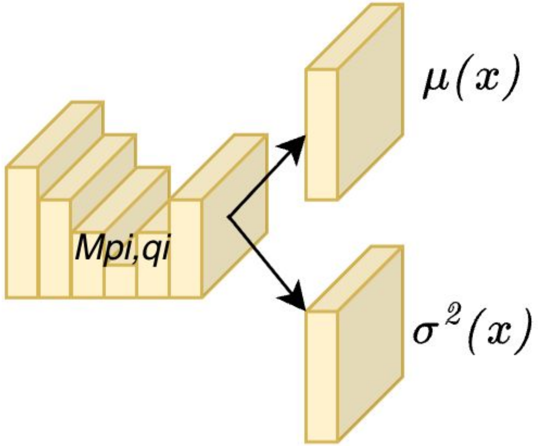


Limitations

Uncertainty

~~Aleatoric~~	***Epistemic***

Only epistemic



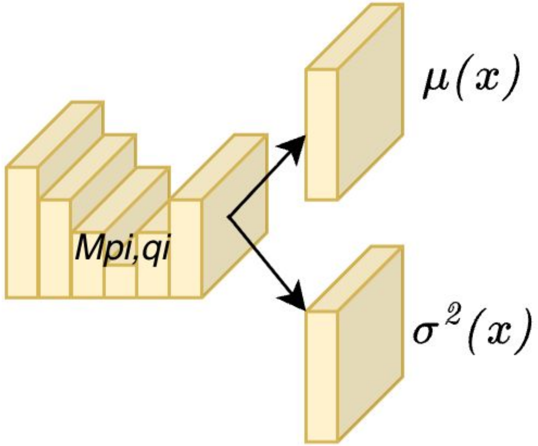
requires likelihood-based loss functions

Limitations

Uncertainty

~~Aleatoric~~	***Epistemic***

Only epistemic



requires likelihood-based loss functions

Dataset

INbreast

Limited patient data to draw conclusions

Dice variance:

- Test set split
- Training seed
- MC Dropout

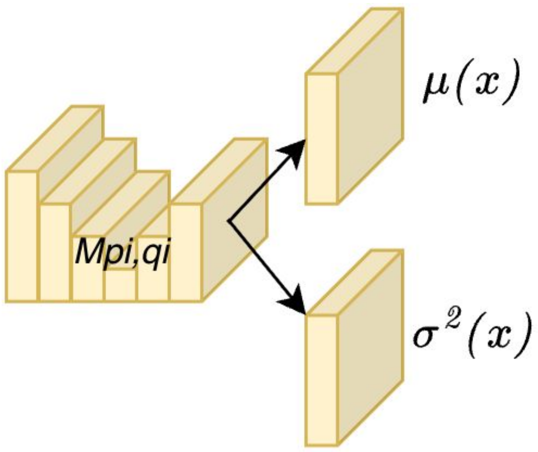
ALL sample variance estimates

Limitations

Uncertainty

~~*Aleatoric*~~	***Epistemic***

Only epistemic



requires likelihood-based loss functions

Dataset

INbreast

Limited patient data to draw conclusions

Dice variance:
- Test set split
- Training seed
- MC Dropout

ALL sample variance estimates

M-SYNTH

Restricted synthetic diversity

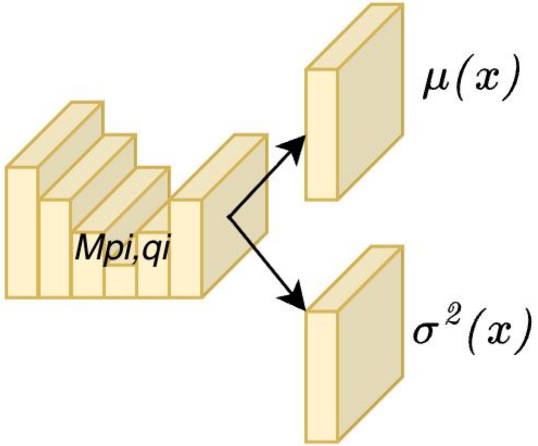
Our configuration:
- Lesion size
- Lesion density
- Breast density

Limitations

Uncertainty

~~*Aleatoric*~~ | ***Epistemic***

Only epistemic



requires likelihood-based loss functions

Dataset

INbreast

Limited patient data to draw conclusions

Dice variance:

- Test set split
- Training seed
- MC Dropout

ALL sample variance estimates

M-SYNTH

Restricted synthetic diversity

Our configuration:

- Lesion size
- Lesion density
- Breast density

Domain gap between patient and synthetic images

Conclusions

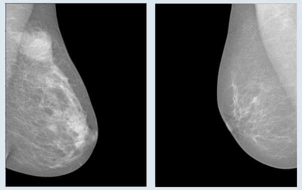
Epistemic

Data

Reader

Case

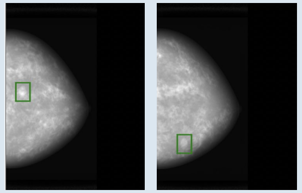
INbreast



publicly available
105 mammograms with lesions
ground truth lesion annotations

$\mathcal{D}_{patient} = (\mathbf{X}_{patient}, \mathbf{Y}_{patient})$

M-SYNTH



publicly available
1200 stochastic knowledged-based models
generated using the VICTRE pipeline
fatty breasts
lesion density of 1.0 and size 5mm
100% of clinically recommended dose

$\mathcal{D}_{synthetic} = (\mathbf{X}_{synthetic}, \mathbf{Y}_{synthetic})$

Context-dependent effects of synthetic data

Optimal ratio requires refinement

Persistent challenging cases still remain

Acknowledgments

U.S. FDA, Center for Devices and Radiological Health, Office of Science and Engineering Laboratories

Department of Radiology, UZ Leuven

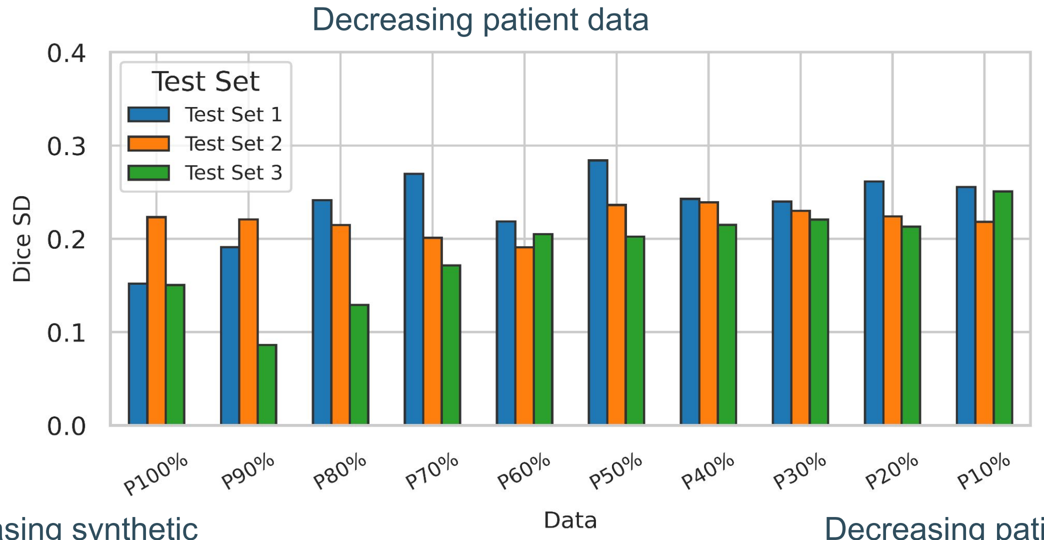
Thank you

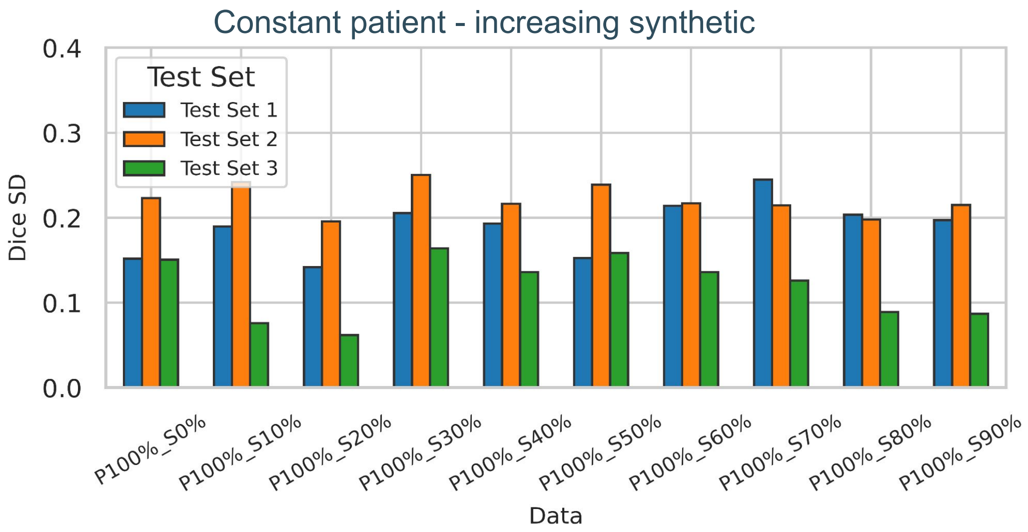
Uncertainty Quantification for Synthetic Medical Images

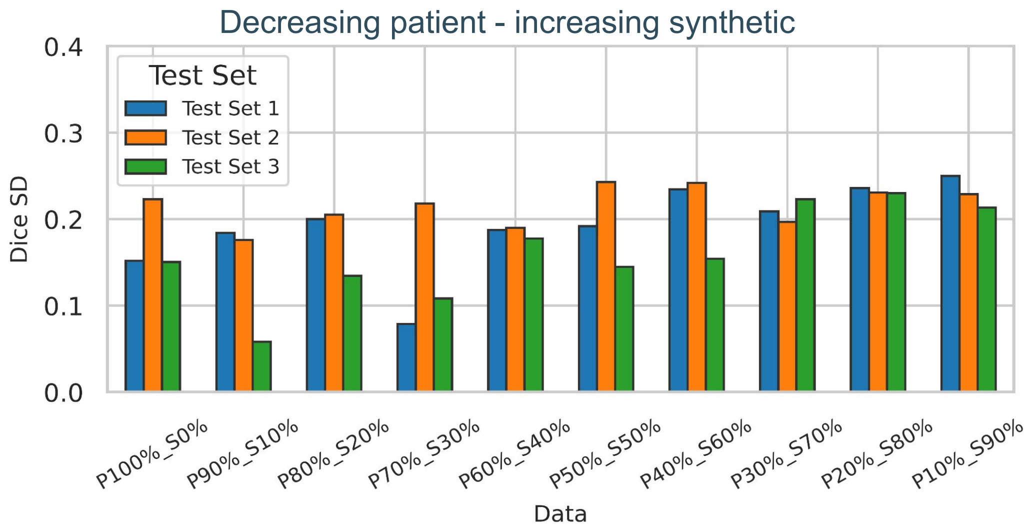
Paper 13928-24 SPIE Medical Imaging 2026

Appendix - Results

Data uncertainty

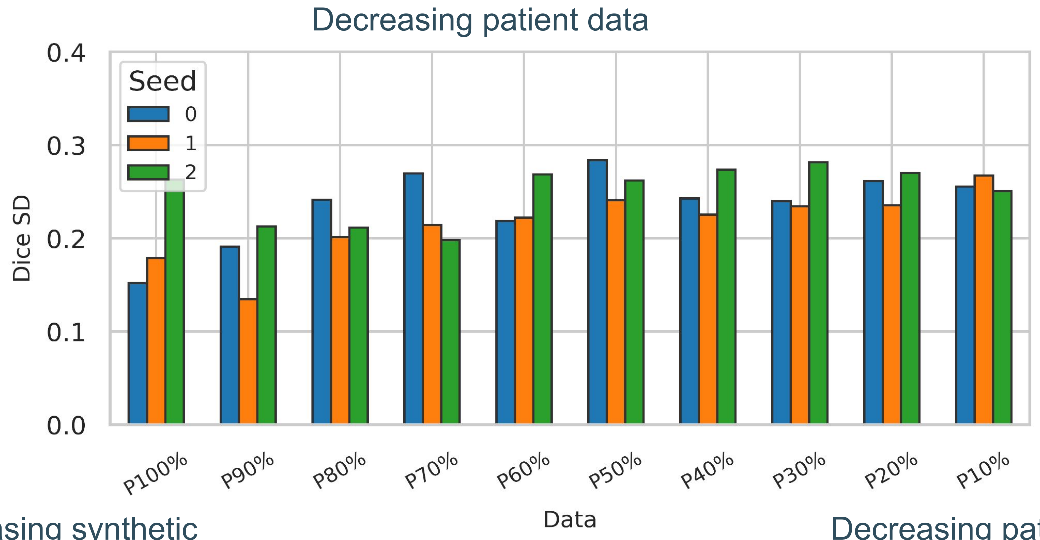


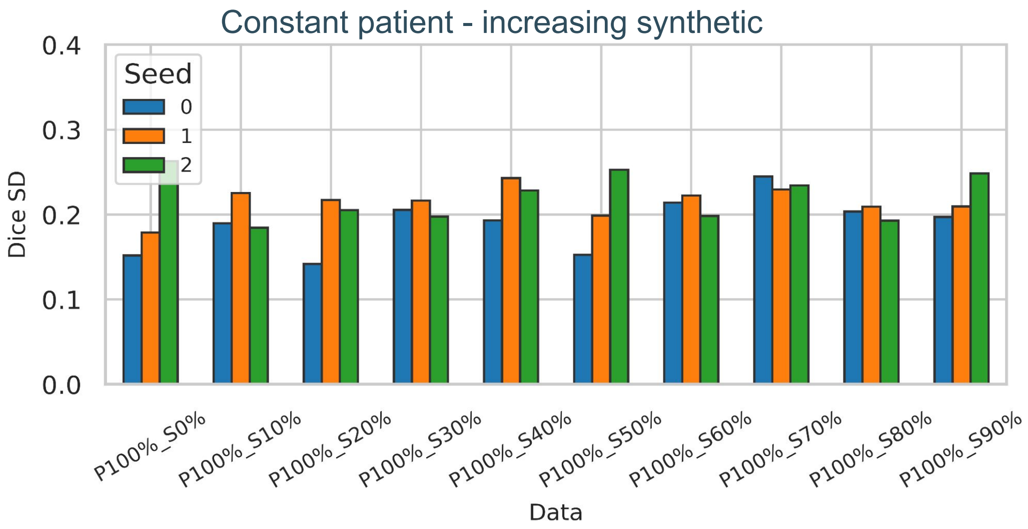


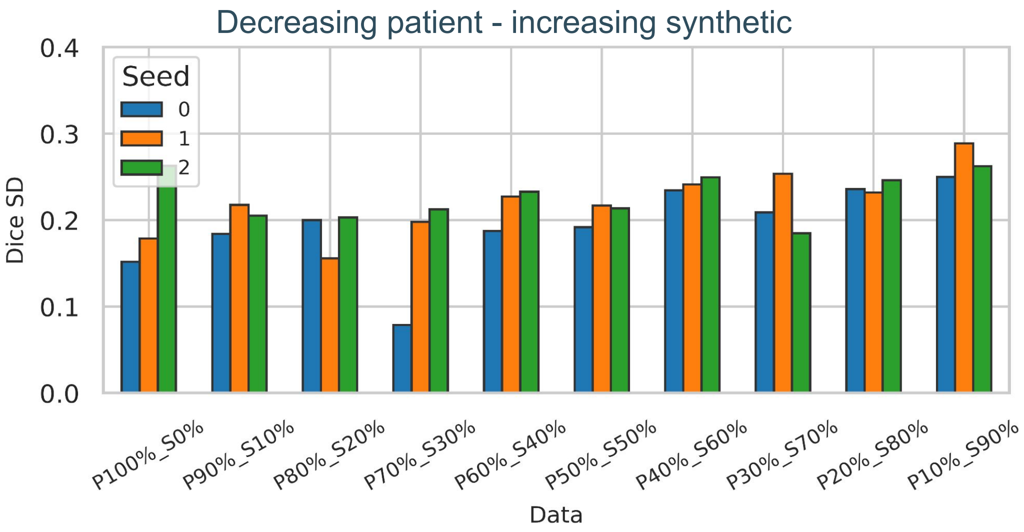


Appendix - Results

Reader uncertainty







Appendix - Results

Case uncertainty