Bootstrapped Physically-Primed Neural Networks for Robust T2 Distribution Estimation in Low-SNR Pancreatic MRI

Hadas Ben-Atya[1] HDS@CAMPUS.TECHNION.AC.IL
Nicole Abramenkov[1] NICOLEA@CAMPUS.TECHNION.AC.IL
Noa Mashiah[1] NOA.MASHIAH@CAMPUS.TECHNION.AC.IL
Luise Brock[1,2] LUISE.BROCK@FAU.DE
Daphna Link-Sourani[1,3] LDAPHNA@BM.TECHNION.AC.IL
Ram Weiss[4] RAMW@RAMBAM.HEALTH.GOV.IL
Moti Freiman[1,3] MOTI.FREIMAN@TECHNION.AC.IL

[1] *Faculty of Biomedical Engineering, Technion – Israel Institute of Technology, Haifa, Israel*
[2] *Pattern Recognition Lab, Friedrich-Alexander-Universität, Erlangen, Germany.*
[3] *The May-Blum-Dahl MRI Research Center, Faculty of Biomedical Engineering, Technion- IIT, Haifa, Israel*
[4] *Department of Pediatrics A, Rambam Healthcare Campus, Haifa, Israel*

Editors: Under Review for MIDL 2026

Abstract

Estimating multi-component T_2 relaxation distribution from Multi-Echo Spin Echo (MESE) MRI is a severely ill-posed inverse problem, traditionally approached with regularized non-negative least squares (NNLS). In abdominal imaging—and in the pancreas in particular—low Signal-to-Noise Ratio (SNR), motion-induced echo corruption, and the extended echo-train structure challenge both classical solvers and deterministic deep learning models. We introduce a **bootstrap-based inference framework for robust distributional T_2 estimation**, which performs stochastic resampling of the echo train and aggregates predictions across multiple echo subsets. This strategy treats the acquisition as a distribution rather than a fixed input, yielding **variance-reduced, physically consistent** estimates and converting deterministic relaxometry networks into **probabilistic ensemble predictors**. Building on the P2T2 architecture, our method applies inference-time bootstrapping to smooth noise-driven artifacts, increase tolerance to corrupted echoes, and enhance fidelity to the underlying relaxation distribution.

We demonstrate a clinical application of the proposed approach for functional and physical assessment of the pancreas. Currently available techniques for noninvasive pancreatic evaluation are limited due to the organ's concealed retroperitoneal location and the procedural risks associated with biopsy, driven in part by the high concentration of proteases that can leak and cause intra-abdominal infection. These constraints highlight the need for functional imaging biomarkers capable of capturing early pathophysiological changes. A prominent example is type 1 diabetes (T1DM), in which progressive destruction of beta cells begins years before overt hyperglycemia, yet no existing imaging modality can assess early inflammation or the decline of pancreatic islets. A further unmet need lies in characterizing pancreatic lesions suspected of malignancy: although malignant and benign lesions differ in their physical properties, current imaging methods do not reliably distinguish between them.

To examine the clinical utility of our method, we evaluate performance in **test–retest reproducibility study** ($N = 7$) and a **T1DM versus healthy differentiation task** ($N = 8$). The proposed approach achieves the lowest Wasserstein distances across repeated scans and demonstrates superior sensitivity to subtle, physiology-driven shifts in the relaxation-time distribution, outperforming classical NNLS and non-bootstrapped deep learning baselines. These results establish inference-time bootstrapping as an effective and practical enhancement for quantitative T_2 relaxometry in low-SNR abdominal imaging, enabling more stable and discriminative estimation of relaxation-time distributions.

Keywords: Multi-component T2, Quantitative MRI, Physics-informed Neural Networks, Bootstrapped Inference, Test-Retest Reliability, Microstructural Biomarkers

1. Introduction

Type 1 Diabetes Mellitus (T1DM) is a chronic autoimmune disease characterized by progressive loss of insulin-producing beta cells in the pancreas. Current diagnostic tools—serological assays and blood glucose measurements—typically identify the disease only after substantial beta-cell destruction has occurred. This motivates the development of non-invasive imaging biomarkers capable of detecting early microstructural alterations, such as inflammation or edema, that precede overt hyperglycemia (Bonner Weir et al., 1983; Jacobsen et al., 2020).

Quantitative Magnetic Resonance Imaging (qMRI), and multi-component T_2 relaxometry, in particular, provide a powerful framework for probing such microscopic changes. Unlike conventional T_2-weighted imaging, which is influenced by acquisition settings and subjective image interpretation, multi-component relaxometry seeks to recover a *distribution* of relaxation times that reflects different water environments within tissue (Radunsky et al., 2021; Fatemi et al., 2020; Margaret Cheng et al., 2012). The MESE signal can be expressed as a Fredholm integral equation of the first kind:

$$y(TE) = \int K(TE, T_2), p(T_2), dT_2 + \eta, \tag{1}$$

where K is the decay kernel—commonly computed via the Extended Phase Graph (EPG) formalism to account for B_1 inhomogeneities (Hennig, 1988)—$p(T_2)$ is the relaxation-time distribution, and η represents measurement noise.

Recovering $p(T_2)$ from discrete MESE measurements is a highly ill-posed inverse problem. Classical methods such as Tikhonov-regularized Non-Negative Least Squares (NNLS) (Doucette et al., 2020; Canales-Rodríguez et al., 2021; Bai et al., 2014; Chatterjee et al., 2018) often yield unstable or overly sparse solutions, especially in abdominal imaging where low SNR and motion corruption disrupt individual echoes. Deep learning approaches, including Model-Informed Machine Learning (MIML) (Yu et al., 2021), accelerate estimation but remain sensitive to echo outliers and may fail to generalize across variable MESE protocols.

In this work, we introduce a **bootstrap-based inference strategy** that transforms deterministic T_2 relaxometry networks into robust probabilistic estimators. At inference time, the method stochastically resamples subsets of the echo train and aggregates the resulting predictions, effectively marginalizing over echo corruption and reducing variance in the recovered distribution $p(T_2)$. This ensemble view treats the acquisition as a distribution rather than a fixed signal, markedly improving the robustness to noise and motion artifacts.

Our approach is based on the Physically Protected T2 (P2T2) architecture (Ben-Atya and Freiman, 2023), which encodes the echo-time schedule directly into the network, allowing generalization across heterogeneous MESE protocols. Although P2T2 mitigates protocol dependence, its deterministic formulation remains vulnerable to aleatoric noise; inference-time bootstrapping resolves this limitation and enhances sensitivity to subtle changes driven by physiology in pancreatic tissue.

In two experimental settings, a **test–retest reproducibility study** ($N = 7$) and a **T1DM versus healthy differentiation task** ($N = 8$), our method achieved the lowest Wasserstein distances and produced the largest distributional separations between-groups. These results demonstrate that bootstrap-enhanced inference substantially improves stability and discriminative power in quantitative T_2 relaxometry for low-SNR abdominal imaging.

Contributions This work makes the following key contributions:

- We introduce an **inference-time bootstrap strategy** for stabilizing multi-component T_2 distribution estimation. By resampling echo subsets and aggregating ensemble predictions, the method reduces variance and increases robustness to low-SNR and motion-corrupted echoes.
- We perform a **comprehensive statistical evaluation** using AUC, Hellinger distance, and the Kolmogorov–Smirnov p-value, demonstrating that bootstrap-enhanced inference yields more reproducible and more discriminative T_2-derived biomarkers than existing classical and deep learning baselines.

2. Methods

2.1. The Inverse Problem and Forward Model

Estimating the T_2 relaxation distribution from multi-echo data can be formulated as an inverse problem, in which a physical signal-decay model is fitted to the acquired MRI echo train. Given multi-echo spin-echo (MESE) data sampled at echo times $TE_{i=1}^{N}$, the objective is to infer a discretized distribution $p(T_2)$ that explains the measured signal $\mathbf{s}$.

This is commonly posed as a regularized least-squares optimization:

$$\widehat{p(T_2)} = \arg\min_{p(T_2)} \sum_{i=1}^{N} |s(TE_i) - s_i|_2^2 + \lambda\mathcal{R}, \tag{2}$$

where s_i is the measured signal, $s(TE_i)$ is the forward-model prediction, and $\mathcal{R}$ enforces smoothness and physical plausibility on the estimated relaxation distribution.

The MRI signal is modeled using the Extended Phase Graph (EPG) formalism (Hennig, 1988; Weigel, 2015). The predicted signal is:

$$s(TE_i) = \sum EPG(TE_i, T_1, T_2, \alpha) \cdot p(T_2) \tag{3}$$

where α denotes the refocusing flip angle train. Thus, the goal is to learn the inverse mapping $f : (\mathbf{s}, \mathbf{TE}) \rightarrow p(T_2)$

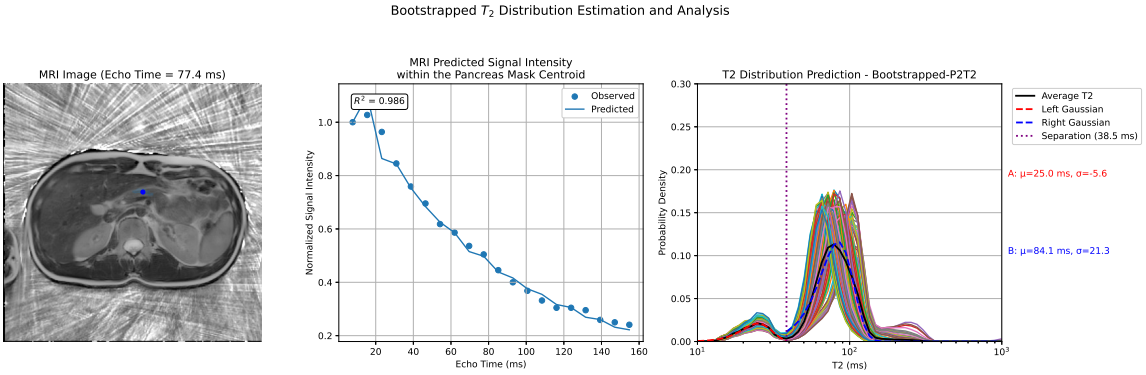


Figure 1: **Representative T_2 distribution estimation workflow (Control Subject). Left:** Axial MRI slice (TE=77.4 ms) with the pancreatic ROI centroid indicated. **Center:** Signal decay analysis: The model accurately fits the physical decay curve ($R^2 = 0.986$), matching the predicted signal (line) to the observed echoes (dots). **Right:** Regional T_2 distribution analysis. Faint colored curves show the final bootstrapped T_2 distributions for individual pixels within the ROI, illustrating local spatial heterogeneity. The solid black curve denotes the spatially averaged distribution across the ROI. Dashed lines depict the Gaussian decomposition of this average into Short-T_2 (red) and Long-T_2 (blue) components.

2.2. Bootstrapped Physically-Primed Networks for Robust T_2 Estimation

Our main contribution is a bootstrap-based reconstruction framework that converts a deterministic T_2 estimator into a robust, ensemble-driven model. Instead of relying on the full echo train, the method repeatedly resamples small subsets of echoes and aggregates the resulting predictions, effectively marginalizing over motion corruption, acquisition noise, and local artifacts. Given an echo train of length N_{total}, the method proceeds as follows:

1. **Subset Resampling:** For each bootstrap iteration $b = 1, \ldots, B$, we randomly sample a subset of $m < N_{\text{total}}$ echoes, yielding reduced signal and TE vectors.

2. **Prediction:** Each subset is processed by a physically-primed model trained for inputs of length m, producing a distribution estimate $\hat{p}_b(T_2)$.

3. **Ensemble Aggregation:** The final T_2 distribution is obtained by averaging bootstrap predictions:

$$\hat{p}_{\text{final}}(T_2) = \frac{1}{B} \sum_{b=1}^{B} \hat{p}_b(T_2).$$

We set $B = 200$ throughout all experiments, which provides an effective balance between computational cost and variance reduction. This ensemble mechanism significantly improves reconstruction stability in low-SNR regimes and reduces susceptibility to motion-corrupted echoes.

2.3. Backbone Architecture

As the underlying predictor, we employ a fully connected network that jointly receives the MRI signal and its corresponding echo-time (TE) schedule. For each voxel, the input vector is defined as

$$\mathbf{x} = [\,\mathbf{s} \oplus \mathbf{TE}\,],$$

where $\mathbf{s}$ denotes the measured signal and $\mathbf{TE}$ the physical echo times. Explicitly encoding TE values—rather than echo indices—allows the model to learn protocol-invariant decay relationships and generalize across heterogeneous acquisition designs.

The backbone consists of 12 fully connected layers with 256 units per layer and ReLU activations. The final layer is a SoftMax classifier over a discretized grid of T_2 values, yielding a normalized relaxation-time distribution. Because the network is physically primed via the TE schedule, it is inherently protocol-agnostic and supports the bootstrap strategy without requiring per-scanner or per-protocol model retraining.

3. Experimental Setup

3.1. Participants and Study Design

Two cohorts were analyzed to evaluate algorithmic stability and clinical sensitivity:

- **Test–Retest Reliability:** Seven healthy volunteers were recruited. Each subject was scanned twice within an interval of approximately 10 minutes to evaluate the repeatability of the estimator in the absence of physiological intervention.
- **Glucose Challenge:** A cohort of eight subjects, comprising four patients with T1DM and four healthy controls. Participants were scanned at two timepoints: baseline (pre-glucose), and after a 30-minute wait (post-glucose intake), to probe metabolic-induced microstructural changes.

All participants provided informed consent, adhering to institutional guidelines.

3.2. Image Acquisition and Segmentation

Abdominal MRI was acquired using a Multi-Echo Spin Echo (MESE) sequence on a 3T scanner. The protocol utilized a train of 32 echoes to sample the decay curve. To mitigate physiological motion artifacts common in abdominal imaging, all echo images were rigidly aligned to the first echo prior to analysis. The pancreas was manually segmented on anatomical images and subdivided into body, and tail regions. These segmentation masks were subsequently resampled and projected onto the co-registered ME-T2W space.

3.3. T2 Distribution Modeling

Figure 2 presents a schematic comparison of the evaluated model architectures. In this study we evaluated three architectures:

1. **MIML (Yu et al., 2021):** A baseline neural network that maps the input signal directly to a T_2 distribution without encoding the echo-time schedule.

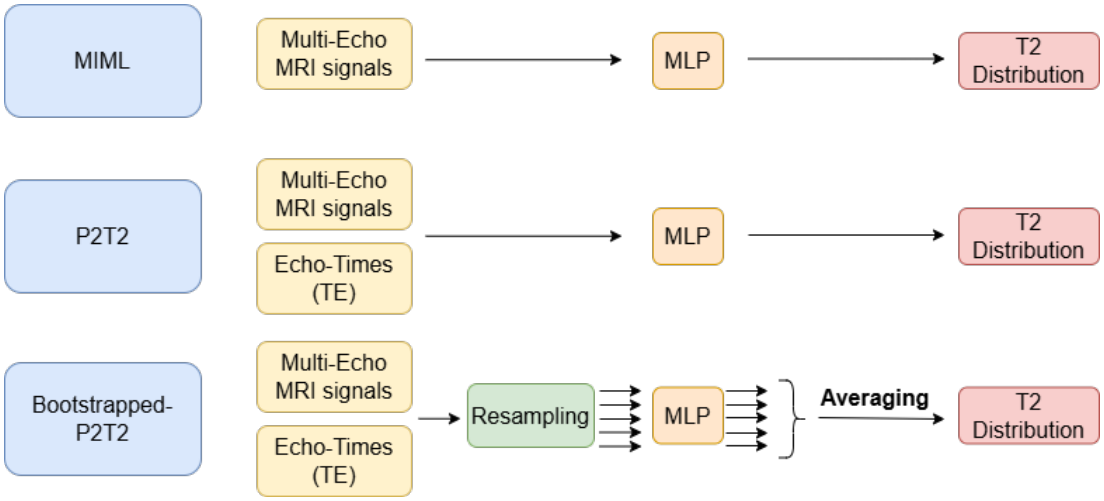


Figure 2: Comparison of the evaluated model architectures. The baseline MIML directly maps signals to T_2 distributions; P2T2 incorporates the explicit TE encoding; and the proposed Bootstrapped-P2T2 adds inference-time resampling and ensemble averaging to improve robustness and stability.

2. P_2T_2 **(Ben-Atya and Freiman, 2023):** An extension of the MIML architecture that concatenates the specific TE acquisition array with the input echo signal, allowing the model to generalize across different protocols.

3. **Bootstrapped-P_2T_2:** Our proposed robust framework, which performs repeated P_2T_2 estimation on random echo subsets ($m < N_{total}$).

Training Data. All networks were trained on synthetic MESE datasets generated via the Extended Phase Graph (EPG) formalism to simulate a wide range of biologically plausible water pools. Each voxel-wise prediction is a discretized probability density function $p_v(T_2)$.

Evaluation. For each anatomical ROI, we compute the Wasserstein distance (W_1) between the *average* distribution from the two scans (test–retest) or between the pre–/post–glucose scans (T1DM vs. control). Lower W_1 indicates better reproducibility; larger between-group differences indicate better sensitivity.

4. Results

4.1. Algorithmic Stability (Test-Retest)

Figure 3 summarizes the reproducibility of the evaluated models across repeated scans, quantified using the Wasserstein distance (W_1) between distributions computed within identical ROIs of the pancreas body and tail. Lower values indicate higher stability. Across all subjects, the proposed Bootstrapped-P2T2 consistently achieved the lowest median and variance of W_1, demonstrating substantially improved robustness in low-SNR abdominal conditions. The deterministic P2T2 model also improved over the baseline MIML approach,

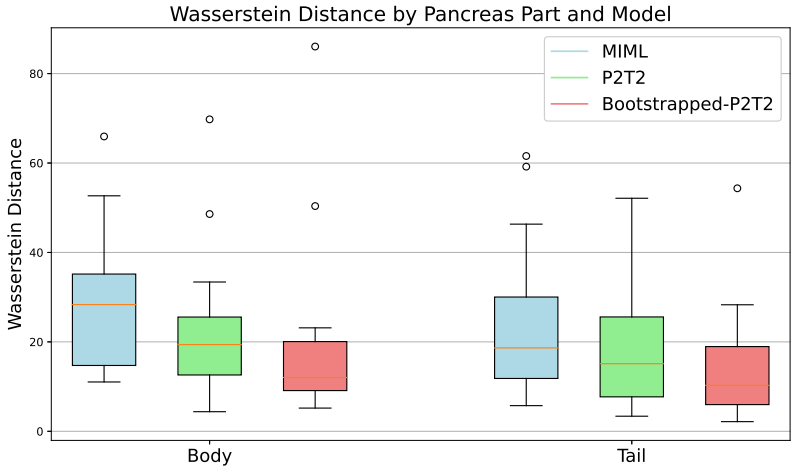
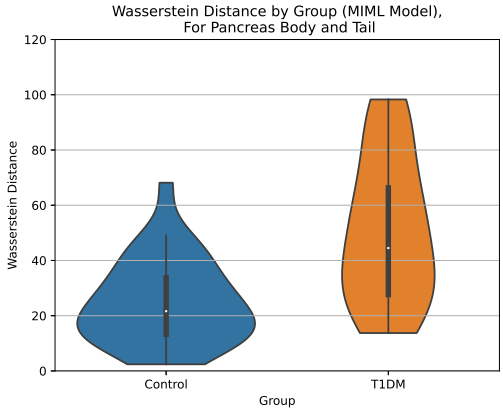


Figure 3: **Test–retest stability across anatomical regions.** Boxplots show the Wasserstein distance (W_1) between repeated scans for the pancreatic body and tail. The proposed **Bootstrapped-P2T2** model (red) exhibits consistently lower median error and reduced variance compared with the deterministic P2T2 (green) and baseline MIML (blue), demonstrating improved robustness and repeatability across both regions.

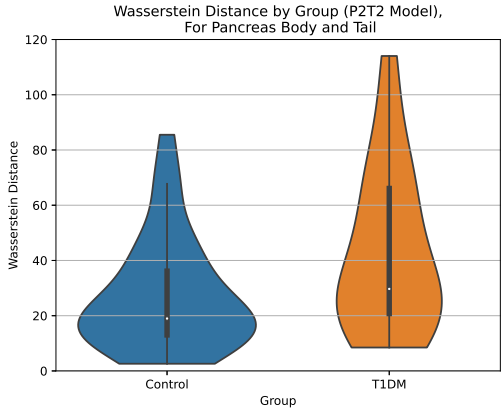
highlighting the advantage of explicitly encoding the TE schedule; however, its variability remained noticeably higher than that of the bootstrapped ensemble. These results indicate that inference-time stochastic resampling plays a critical role in stabilizing multi-echo T_2 estimates, yielding markedly more reproducible distributions in the anatomical regions most relevant to pancreatic endocrine tissue.

4.2. Separation Between T1DM and Controls

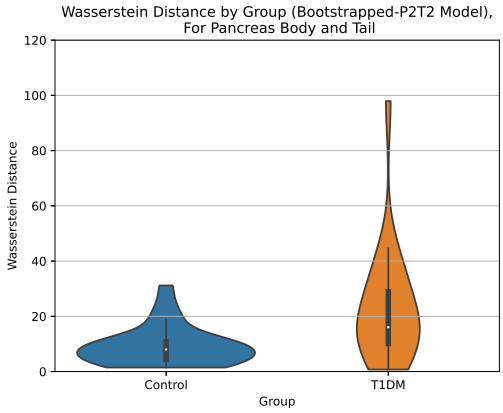
We next assessed each model's ability to detect physiological changes between the pre- and post-glucose scans. Table 4.2 summarizes the statistical analysis results. Figure 4 (a-c) shows the distribution of Wasserstein distances (W_1) between timepoints for each subject in the pancreas body and tail. Bootstrapped-P2T2 provided the clearest group separation, with T1DM subjects exhibiting markedly larger distributional shifts than healthy controls and with minimal overlap between groups (Fig. 4c). Deterministic P2T2 captured the same directional trend but with higher overlap (Fig. 4b), reflecting its greater sensitivity to noise. MIML showed statistical differences but substantial overlap between groups (Fig. 4a), indicating that much of its variation is not physiologically driven. Classical T_2 scalar reconstruction, including ΔT_2 and ΔM_0, showed even higher overlap (Fig. 4d–e), demonstrating their limited ability to capture microstructural alterations. Figure 5 presents two representative subjects, illustrating the markedly larger distributional shift seen in a T1DM participant relative to the minimal shift in a healthy control.



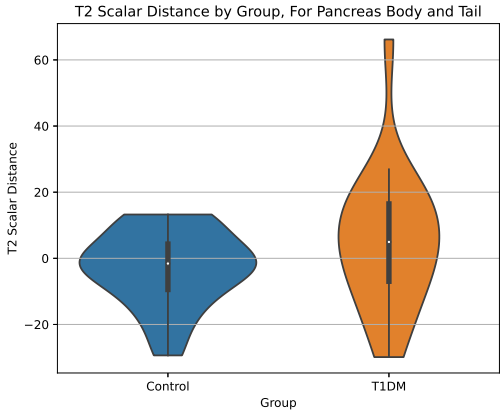
(a) MIML



(b) P2T2



(c) Bootstrapped-P2T2



(d) Scalar T_2 (ΔT_2)

(e) Scalar M_0 (ΔM_0)

Figure 4: **Sensitivity analysis across models (Pancreas Body & Tail).** (a–c) Distribution-based deep learning models. The proposed Bootstrapped-P2T2 (c) demonstrates the clearest group separation between T1DM and Controls. (d–e) Classical scalar metrics show larger overlap, highlighting the improved sensitivity gained from full-distribution modeling.

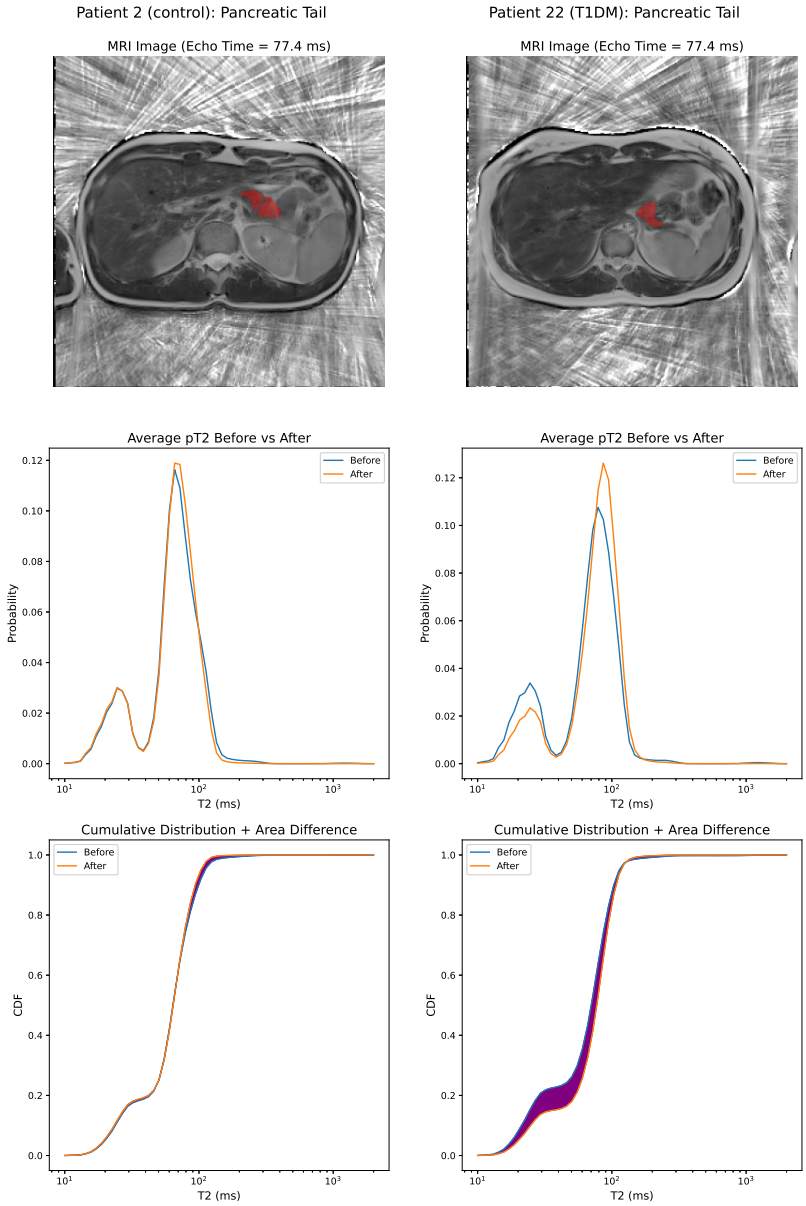


Figure 5: Representative response to glucose challenge in the pancreatic tail, using the Bootstrapped-P2T2. **Left:** Healthy Control subject. **Right:** T1DM subject. **Top Row:** Axial MRI (10th echo) with the pancreatic tail segmentation overlaid in red. **Middle Row:** Average T_2 distributions ($\bar{p}(T_2)$) within the ROI before (blue) and after (orange) glucose intake. **Bottom Row:** Cumulative Distribution Functions (CDFs) highlighting the gap between timepoints. Note the significantly larger distributional shift in the T1DM subject compared to the stable profile of the control, indicating an altered microstructural response to metabolic stress.

Model	AUC ↑	Hellinger ↑	KS p-value ↓
MIML	**0.823**	0.287	2.36×10^{-3}
P2T2-FC	0.693	0.237	114.0×10^{-3}
Bootstrapped-P2T2	0.800	**0.427**	$\mathbf{0.47 \times 10^{-3}}$
T2 Scalar (ΔT2)	0.61	0.254	202×10^{-3}
T2 Scalar (ΔM0)	0.605	0.180	168×10^{-3}

Table 1: Comparison of separation metrics across models, within the Pancreas Body and Tail.

5. Discussion

This study demonstrates that inference-time bootstrapping substantially improves the robustness and sensitivity of multi-component T_2 relaxometry in abdominal MRI. Across both experiments, test–retest reproducibility and the glucose-challenge paradigm, the proposed Bootstrapped-P2T2 model consistently outperformed deterministic baselines. In the reproducibility analysis, Bootstrapped-P2T2 achieved the lowest and most stable Wasserstein distances, revealing a clear hierarchy of performance (Bootstrapped-P2T2 > P2T2 > MIML) and indicating that echo-wise stochastic resampling effectively suppresses noise-amplifying artifacts inherent to MESE acquisitions while the physics-based TE encoding retains protocol invariance. In the glucose-challenge cohort, Bootstrapped-P2T2 detected the strongest and most physiologically plausible distributional shifts between pre- and post-glucose scans, in the pancreas body and tail, regions known to reflect endocrine tissue function; deterministic P2T2 captured similar but weaker trends, whereas scalar T_2 mapping showed limited discriminative ability. While this work is limited by modest sample sizes and by fixed choices of echo subsampling and bootstrap count, the results suggest that inference-time bootstrapping stabilizes the ill-posed T_2 inversion problem and enhances sensitivity to subtle microstructural changes. Taken together, these findings position Bootstrapped-P2T2 as a robust and biologically meaningful tool for quantitative relaxometry, with potential relevance for developing non-invasive markers of early pancreatic dysfunction in T1DM.

6. Conclusion

We presented a novel **Bootstrapped-P2T2 framework** for multi-component T_2 relaxometry that integrates physics-informed neural networks with inference-time bootstrapping to enhance robustness and distributional fidelity. This approach improves test–retest reproducibility and provides the clearest separation between T1DM and healthy pancreatic tissue among the evaluated methods. Because no reliable noninvasive tools currently exist for early functional assessment of the pancreas, the ability to recover stable, physiologically meaningful relaxation-time distributions remains a critical unmet need. By mitigating low-SNR and motion-related artifacts while capturing subtle microstructural changes, Bootstrapped-P2T2 directly addresses this gap. Overall, the method represents a promising step toward **clinically actionable quantitative pancreatic MRI**, with potential applications in early T1DM detection and pancreatic lesion characterization.

Acknowledgments

This study was supported in part by a research grant from the Technion's EVPR Foundation for Collaborative Research with the Rambam Healthcare Campus Researchers.

References

Ruiliang Bai, Cheng Guan Koay, Elizabeth Hutchinson, and Peter J Basser. A framework for accurate determination of the t2 distribution from multiple echo magnitude mri images. *Journal of Magnetic Resonance*, 244:53–63, 2014.

Hadas Ben-Atya and Moti Freiman. P2T2: A physically-primed deep-neural-network approach for robust T2 distribution estimation from quantitative T2-weighted MRI. *Computerized Medical Imaging and Graphics*, page 102240, 5 2023. ISSN 0895-6111. doi: 10.1016/J.COMPMEDIMAG.2023.102240. URL https://linkinghub.elsevier.com/retrieve/pii/S0895611123000587.

S. Bonner Weir, D. F. Trent, and G. C. Weir. Partial pancreatectomy in the rat and subsequent defect in glucose-induced insulin release. *The Journal of Clinical Investigation*, 71(6):1544–1553, 6 1983. ISSN 0021-9738. doi: 10.1172/JCI110910.

Erick Jorge Canales-Rodríguez, Marco Pizzolato, Gian Franco Piredda, Tom Hilbert, Nicolas Kunz, Caroline Pot, Thomas Yu, Raymond Salvador, Edith Pomarol-Clotet, Tobias Kober, et al. Comparison of non-parametric t2 relaxometry methods for myelin water quantification. *Medical Image Analysis*, 69:101959, 2021.

Sudhanya Chatterjee, Olivier Commowick, Onur Afacan, Simon K Warfield, and Christian Barillot. Multi-compartment model of brain tissues from t2 relaxometry mri using gamma distribution. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 141–144. IEEE, 2018.

Jonathan Doucette, Christian Kames, and Alexander Rauscher. Decaes–decomposition and component analysis of exponential signals. *Zeitschrift für Medizinische Physik*, 30(4): 271–278, 2020.

Yaghoub Fatemi, Habibollah Danyali, Mohammad Sadegh Helfroush, and Houshang Amiri. Fast T2 mapping using multi-echo spin-echo MRI: A linear order approach. *Magnetic Resonance in Medicine*, 84(5):2815–2830, 11 2020. ISSN 1522-2594. doi: 10.1002/MRM.28309. URL https://doi.org/10.1002/mrm.28309.

Juergen Hennig. Multiecho imaging sequences with low refocusing flip angles. *Journal of Magnetic Resonance (1969)*, 78(3):397–407, 1988.

Laura M. Jacobsen, Laura Bocchino, Carmella Evans-Molina, Linda DiMeglio, Robin Goland, Darrell M. Wilson, Mark A. Atkinson, Tandy Aye, William E. Russell, John M. Wentworth, David Boulware, Susan Geyer, and Jay M. Sosenko. The risk of progression to type 1 diabetes is highly variable in individuals with multiple autoantibodies following screening. *Diabetologia*, 63(3):588, 3 2020. ISSN 14320428. doi: 10.1007/

S00125-019-05047-W. URL /pmc/articles/PMC7229995//pmc/articles/PMC7229995/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7229995/.

Hai-Ling Margaret Cheng, Nikola Stikov, Nilesh R Ghugre, and Graham A Wright. Practical medical applications of quantitative mr relaxometry. *Journal of Magnetic Resonance Imaging*, 36(4):805–824, 2012.

Dvir Radunsky, Neta Stern, Jannette Nassar, Galia Tsarfaty, Tamar Blumenfeld-Katzir, and Noam Ben-Eliezer. Quantitative platform for accurate and reproducible assessment of transverse (T2) relaxation time. *NMR in Biomedicine*, 34(8), 8 2021. ISSN 10991492. doi: 10.1002/NBM.4537/FORMAT/PDF. URL https://pubmed.ncbi.nlm.nih.gov/33993573/.

Matthias Weigel. Extended phase graphs: dephasing, rf pulses, and echoes-pure and simple. *Journal of Magnetic Resonance Imaging*, 41(2):266–295, 2015.

Thomas Yu, Erick Jorge Canales-Rodríguez, Marco Pizzolato, Gian Franco Piredda, Tom Hilbert, Elda Fischi-Gomez, Matthias Weigel, Muhamed Barakovic, Meritxell Bach Cuadra, Cristina Granziera, et al. Model-informed machine learning for multi-component t2 relaxometry. *Medical Image Analysis*, 69:101940, 2021.