Supplementary Information

Supplementary Methods

Network Input Preparation

Subject-level clinical and acquisition variables were preprocessed for use as conditioning inputs to the model's FiLM layers. The following features were included: age, sex, TSPO genotype (HAB/MAB), diagnostic group, scanner type, injected dose, body weight, tracer identity, and visit number. Continuous variables were standardised, and categorical variables were ordinal-encoded and used as integer indices for embedding layers. All variables were normalised on the training set and the same transformation was applied to test partitions.

PET volumes were standardised (z-scored) and, to enable uniform 3D patch extraction, each volume was zero-padded to a power-of-two cubic dimension based on the largest spatial axis. To limit the possibility that the network would become insensitive to small errors in the late tail, which could nonetheless affect V_T estimates, plasma curves were standardised using separate z-scoring for early and late time segments, defined heuristically as greater or less than 900 seconds. This threshold corresponds to the average post-peak inflection point and was fixed across experiments. Standardisation was weighted by the frame length of each time point.

Image-derived input function

The IDIF was computed following the approach proposed by Maccioni et al.[14], using the carotid siphons as the extraction region. High-intensity voxels were automatically selected from early PET frames (0–2 min), refined using voxel-wise correlation filtering, and a tri-exponential fit was applied to denoise the signal. To obtain a fully corrected plasma input, genotype-specific population-based correction curves were used. These curves, derived from King's College London data, were stratified by HAB and MAB groups. When applying these correction curves to Dataset 1 (King's College London), a leave-one-out strategy was used such that, if the evaluated subject was included in the corresponding genotype-specific population set, their curve was excluded from the population average. Plasma activity was computed using the following correction equations:

$$\begin{cases} C_{\text{plasma-uncorrected}} = \text{POB} \cdot C_{\text{blood}} \\ C_{\text{plasma-corrected}} = \text{PPF} \cdot C_{\text{plasma-uncorrected}} \end{cases} \tag{1}$$

where POB and PPF represent population-averaged parent fraction and plasma-to-blood ratio curves, respectively. C_{blood} is the whole blood tracer concentration, $C_{plasma-uncorrected}$ and $C_{plasma-corrected}$ are plasma tracer concentration before and after correction for radiometabolites.

Compartmental modelling:

To further characterise performance, full kinetic modelling was conducted using a 2T4K[39]. The mathematical formulation of the model is provided by the following equations:

$$\begin{cases} \dot{C}_1(t) = K_1 C_p(t) - (k_2 + k_3) C_1(t) + k_4 C_2(t), & C_1(0) = 0 \\ \dot{C}_2(t) = k_3 C_1(t) - k_4 C_2(t), & C_2(0) = 0 \\ C(t) = V_b C_b + (1 - V_b)(C_1(t) + C_2(t)) \end{cases} \tag{2}$$

where C_1 and C_2 are tracer activity in the two tissue compartments, C_p is the model input function, C_b is the whole blood tracer concentration, K_1, k_2, k_3, k_4, and V_b are the model microparameters[40]. Specifically, microparameter estimation was performed by fitting the model on regional TACs using a weighted non-linear least-square estimator with weights chosen optimally as suggested by the literature[41]. Both C_p or C_b input functions were corrected for delay (i.e., the variable time of appearance of tracer radioactivity in blood depending on the time required for the tracer to travel from the arterial sampling site to the region of interest[42].

For each input function class, models were fit under three input configurations:

1. **TRUE:** $C_p = C_{\text{plasma-corrected (true)}}, \quad C_b = C_{\text{blood (true)}}$
2. **NN-unBlood:** $C_p = C_{\text{plasma-predicted (NN-unBlood)}}, \quad C_b = C_{\text{blood (true)}}$
3. **NN-dPET:** $C_p = C_{\text{plasma-predicted (dPET)}}, \quad C_b = C_{\text{plasma-predicted (dPET)}}$

In case of Dataset 2, for which the true C_{blood} was not available, the true $C_{plasma-uncorrected}$ was used instead.

In all cases, a 1T1K estimation of K_1 parameter[14], was also implemented for comparison. Fits where the resulting V_T exceeded 15 were excluded and considered indicative of numerical instability. These cases accounted for approximately 9% of V_T values across all ROIs.

Supplementary Discussion

The following supplementary notes provide additional context and methodological considerations beyond the main Discussion.

The datasets analysed in this work were reconstructed with different pipelines, including distinct approaches to motion correction. These differences reflect the heterogeneity typically encountered in multi-centre PET studies and underscore the importance of methods that can generalise across acquisition settings.

In this first implementation of ASTRA, minimal preprocessing was deliberately applied to avoid potential quantification distortions; nevertheless, future work could explore such operations to assess their impact on robustness and generalisability.

Estimation of the AIF from PET images alone (NN-dPET) also remains affected by reconstruction and temporal framing. In particular, misestimation of the early peak due to coarse framing or smoothing can propagate into the inferred input function. Although the present framework demonstrated robustness to these factors, as evidenced by high correlation and coefficient of determination during cross-validation, further mitigation may be possible. Direct use of listmode data, bypassing reconstruction, could enhance fidelity in capturing the rapid tracer kinetics around the peak.

In Dataset 3, it was examined whether previously reported SUVR group differences could be reproduced with AI-derived blood metrics in the absence of true arterial data. Dataset 3 consisted of a subset of subjects from published findings[26] that did not have arterial blood data and excluded individuals already present in Dataset 1. ASTRA-derived DVR estimates were found to closely parallel the SUVR results, confirming that the method can recover clinically meaningful contrasts without invasive input functions. An exception was observed in the pituitary, where neither SUVR nor DVR showed significant effects, likely reflecting the noise sensitivity of small ROIs and the limited cohort size. V_T metrics were also explored, but no significant group differences were detected. It is possible that inter-subject variability, compounded by method-estimation errors, increased dispersion and reduced the ability to detect group-level effects. Further clarification of the sensitivity of ASTRA-derived V_T will require larger cohorts with arterial sampling as reference standard.

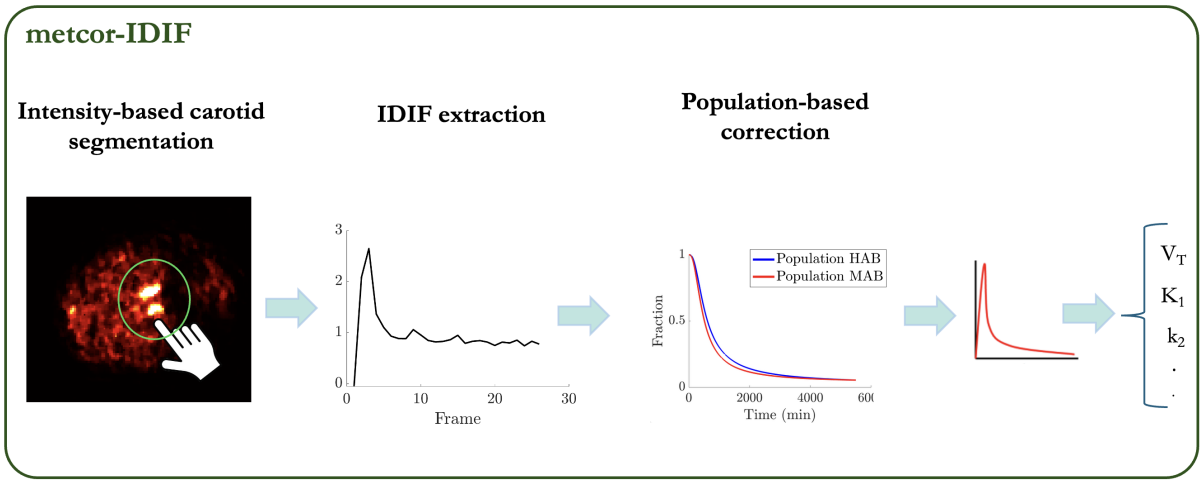
Performance in this study was evaluated against blood-derived input functions taken as the true reference. However, arterial sampling is itself prone to dispersion, delay, calibration, and metabolite-correction errors, and thus does not represent an absolute ground truth. Comparisons against blood data are therefore partly circular, as discrepancies may reflect imperfections in the reference rather than in ASTRA estimates. Future work including repeat blood sampling may help to disentangle these effects.

In the current implementation, regularisation was weighted by frame duration to improve robustness in the tail of the signal, which may have come at the cost of reduced fidelity around the peak as reflected in microparameter and curve estimates, a limitation that could be addressed in future work with larger datasets.

Physics-informed implementations of ASTRA[43] could embed prior physiological knowledge, which may be particularly valuable for transfer learning across sites or tracers. In parallel, saliency maps or analyses of hidden-layer embeddings could improve interpretability by clarifying which image features drive AIF estimation and how these relate to physiology.

Finally, the degree of compression in the autoencoder offers another direction: while a small latent dimensionality was adopted here due to memory constraints, future tests on more powerful machines could systematically vary this capacity to determine the optimal representation size, which may be larger than the one used here. This would clarify the efficiency–robustness trade-off and guide the design of new models for broader application.

Supplementary Figures



Supplementary Figure 1. Metcor-IDIF pipeline. The IDIF is extracted from the segmented carotid region. A population-based correction is applied using reference curves for different subject groups (HAB and MAB), yielding a metabolite-corrected IDIF for subsequent analysis.

892 *Supplementary Tables*

Supplementary Table 1. Participant demographics and diagnostic composition across datasets. **Legend:** SZ = Schizophrenia, UHR = Ultra-High Risk of Psychosis, cLBP = chronic Low Back Pain, KOA = Knee Osteo-Arthritis, HC = Healthy Controls, M/F = Male/Female, HAB/MAB = High/Mixed-Affinity Binder.

Dataset	N	HC	SZ	UHR	cLBP	KOA	Age (y)	M/F	HAB/MAB	Re-scans
1	90	66	12	12	–	–	32.5 ± 13.3	62/28	64/26	0
2	52	10	–	–	27	15	55.6 ± 16.0	27/25	32/20	13
3	32	11	–	–	–	21	63.5 ± 10.4	15/17	19/13	0
4	5	–	5	–	–	–	49.4 ± 9.4	5/0	5/0	0

Supplementary Table 2. **Key hyperparameters for each ASTRA module.** Legend: lr = learning rate; wd = weight decay.

Module	Hyperparameter / Setting
M1	Patch size / stride: 32^3 voxels / 16 voxels Conv3D filters: {16, 32, 64, 128, 256, 512} Conv layers per block (incl. bottleneck): {2, 2, 2, 2, 2, 2, 2} FiLM layers: 2 Latent size: 128 (per patch) Activations: GELU (dense), Swish (conv) Dropout rate: 10% Gaussian noise: $\sigma = 0.01$ Optimizer: Adam (lr = 1×10^{-4}) Loss: Mean Squared Error (MSE) on reconstructed image patches Regularisation: latent mean/variance normalisation; latent decorrelation loss; FiLM commitment regularisation; MSE on Laplacian of reconstructed patches
M2	Conv3D filters: {32, 32, 32, 32, 32} Conv layers per block: {2, 2, 2, 2, 2} Dense units per block: {16} Dense layers per block: {2} Attention heads: 2 Attention dense units: 64 Activations: Swish (conv), GELU (dense) Conv FiLM dense units: 128 Conv FiLM layers: 2 Dense FiLM dense units: {64, 32, 16, 8, 4, 2} Dense FiLM layers: {2, 2, 2, 2, 2} Dropout rate: 10% Gaussian noise: $\sigma = 0.01$ Optimizer: AdamW (lr = 1×10^{-4}, Conv wd = 1×10^{-2}, Dense wd = 1×10^{-1}) Loss: MSE on predicted plasma signal Regularisation: MSE on Laplacian of predicted plasma signal; FiLM commitment regularisation
M3	Dense units per block: {64} Dense layers per block: {2} Attention heads: 4 Attention dense units: 64 FiLM dense units: {128, 64, 32, 16, 8, 4, 2} FiLM layers: {2, 2, 2, 2, 2, 2, 2} Activation: GELU Dropout rate: 50% Gaussian noise: $\sigma = 0.01$ (input), $\sigma = 0.1$ (dense/latent) Optimizer: AdamW (lr = 1×10^{-4}, wd = 1×10^{-1}) Loss: Gaussian Negative Log-Likelihood (NLL) on predicted plasma signal Regularisation: MSE on Laplacian of predicted plasma signal; FiLM commitment regularisation

Supplementary Table 3. Component-wise parameter counts for modules M1–M3, based on Keras model summaries.

Module	Component	Parameters
M1	Embeddings + FiLM modulation	∼2K
	Conv3D Encoder Stack	∼16.0M
	Latent Projection	∼140K
	Conv3D Decoder Stack	∼16.0M
	Total	**32.1M**
M2	Embeddings + FiLM modulation	∼8K
	Temporal Conv3D	∼30.8M
	Transformer Encoder Blocks	∼56K
	Total	**∼30.9M**
M3	Embeddings + FiLM modulation	∼8K
	Transformer Encoder Blocks	∼205K
	Output Heads (mean & std)	∼21K
	Total	**∼234K**

Supplementary Table 4. Dataset 1, NN-unBlood: V_T regression metrics across tuning and cross-validation folds. Metrics are averaged over all ROIs and test subjects.

Fold	MAE	RMSE	R^2	CCC	N
Tuning fold					
0	0.341	0.471	0.972	0.921	9
Cross-validation folds					
1	0.464	0.610	0.892	0.917	9
2	0.353	0.477	0.923	0.933	9
3	0.457	0.528	0.889	0.942	9
4	0.378	0.529	0.917	0.952	9
5	0.348	0.491	0.968	0.939	9
6	0.751	1.067	0.785	0.841	9
7	0.462	0.523	0.879	0.927	9
8	0.486	0.598	0.823	0.898	9
9	0.498	0.608	0.816	0.898	9
Mean ± SD (CV)	0.466 ± 0.121	0.603 ± 0.181	0.877 ± 0.059	0.916 ± 0.034	81

Supplementary Table 5. Dataset 1, NN-dPET: V_T regression metrics across tuning and cross-validation folds. Metrics are averaged over all ROIs and test subjects.

Fold	MAE	RMSE	R^2	CCC	N
Tuning fold					
0	0.666	0.869	0.861	0.771	9
Cross-validation folds					
1	0.620	0.759	0.778	0.870	9
2	0.475	0.547	0.842	0.916	9
3	0.652	0.731	0.796	0.874	9
4	0.733	0.875	0.822	0.828	9
5	0.697	0.844	0.762	0.828	9
6	0.871	1.155	0.600	0.690	9
7	0.583	0.682	0.723	0.844	9
8	0.664	0.747	0.646	0.776	9
9	0.661	0.744	0.711	0.806	9
Mean ± SD (CV)	0.662 ± 0.108	0.787 ± 0.167	0.742 ± 0.080	0.826 ± 0.065	81

Supplementary Table 6. Dataset 2, NN-unBlood: V_T regression metrics across tuning and cross-validation folds. Metrics are averaged over all ROIs and test subjects.

Fold	MAE	RMSE	R^2	CCC	N
Tuning fold					
0	0.583	0.640	0.973	0.925	5
Cross-validation folds					
1	0.357	0.424	0.844	0.767	5
2	0.342	0.456	0.687	0.824	5
3	0.200	0.279	0.942	0.934	5
4	0.292	0.408	0.865	0.856	5
5	0.231	0.285	0.746	0.841	5
6	0.740	0.934	0.709	0.529	5
7	0.358	0.387	0.984	0.942	5
8	0.379	0.476	0.312	0.552	5
9	0.222	0.227	0.948	0.957	5
10	0.503	0.509	0.952	0.490	2
Mean ± SD (CV)	0.362 ± 0.160	0.439 ± 0.197	0.799 ± 0.202	0.769 ± 0.180	47

Supplementary Table 7. Dataset 2, NN-dPET: V_T regression metrics across tuning and cross-validation folds

Fold	MAE	RMSE	R^2	CCC	N
Tuning fold					
0	0.783	0.920	0.836	0.735	5
Cross-validation folds					
1	0.401	0.439	0.555	0.686	5
2	0.440	0.599	0.670	0.780	5
3	0.337	0.425	0.945	0.936	5
4	0.384	0.522	0.934	0.763	5
5	0.384	0.433	0.488	0.694	5
6	0.777	0.921	0.924	0.533	5
7	0.337	0.425	0.945	0.936	5
8	0.533	0.738	0.057	0.224	5
9	0.499	0.543	0.566	0.694	5
10	0.609	0.704	0.913	0.481	2
Mean ± SD (CV)	0.470 ± 0.139	0.575 ± 0.167	0.700 ± 0.292	0.673 ± 0.215	47

Supplementary Table 8. Comparison of parameter estimates for Dataset 1. Values are median ± Median Absolute Deviation (MAD). Inter-subject correlation (ρ) and concordance (CCC) metrics are reported. Abbreviations: 1T1K implemented as in[14]

Parameter	Method	Median ± MAD	Inter-subject ρ	Inter-subject CCC
K_1	True	0.139 ± 0.031	–	–
	NN-unBlood	0.145 ± 0.029	0.961 ± 0.004	0.959 ± 0.004
	p-metcorBlood	0.120 ± 0.026	0.987 ± 0.002	0.875 ± 0.007
	NN-dPET	0.138 ± 0.020	0.576 ± 0.019	0.564 ± 0.021
	p-metcorIDIF	0.607 ± 0.123	0.615 ± 0.026	0.046 ± 0.004
K_1-1T1K	True	0.130 ± 0.026	–	–
	NN-unBlood	0.130 ± 0.028	0.993 ± 0.001	0.991 ± 0.001
	p-metcorBlood	0.107 ± 0.021	0.990 ± 0.000	0.822 ± 0.008
	NN-dPET	0.126 ± 0.018	0.660 ± 0.018	0.652 ± 0.017
	p-metcorIDIF	0.308 ± 0.039	0.757 ± 0.020	0.104 ± 0.009
k_2	True	0.128 ± 0.014	–	–
	NN-unBlood	0.135 ± 0.025	0.481 ± 0.082	0.436 ± 0.095
	p-metcorBlood	0.146 ± 0.017	0.760 ± 0.058	0.681 ± 0.076
	NN-dPET	0.130 ± 0.020	0.478 ± 0.099	0.453 ± 0.104
	p-metcorIDIF	0.512 ± 0.085	0.313 ± 0.065	0.021 ± 0.009
k_3	True	0.073 ± 0.016	–	–
	NN-unBlood	0.070 ± 0.019	0.644 ± 0.051	0.539 ± 0.058
	p-metcorBlood	0.064 ± 0.018	0.678 ± 0.071	0.542 ± 0.075
	NN-dPET	0.073 ± 0.021	0.657 ± 0.064	0.620 ± 0.070
	p-metcorIDIF	0.069 ± 0.011	0.653 ± 0.048	0.602 ± 0.048
k_4	True	0.025 ± 0.003	–	–
	NN-unBlood	0.024 ± 0.005	0.226 ± 0.036	0.064 ± 0.022
	p-metcorBlood	0.026 ± 0.004	0.833 ± 0.035	0.751 ± 0.045
	NN-dPET	0.024 ± 0.005	0.701 ± 0.048	0.677 ± 0.059
	p-metcorIDIF	0.026 ± 0.004	0.499 ± 0.112	0.485 ± 0.115
V_b	True	0.070 ± 0.008	–	–
	NN-unBlood	0.068 ± 0.008	0.930 ± 0.014	0.924 ± 0.015
	p-metcorBlood	0.068 ± 0.008	0.982 ± 0.006	0.975 ± 0.008
	NN-dPET	0.060 ± 0.009	0.535 ± 0.074	0.468 ± 0.078
	p-metcorIDIF	0.135 ± 0.024	0.318 ± 0.158	0.058 ± 0.034
V_T	True	4.293 ± 0.804	–	–
	NN-unBlood	4.115 ± 0.961	0.882 ± 0.025	0.877 ± 0.027
	p-metcorBlood	3.180 ± 0.714	0.839 ± 0.017	0.681 ± 0.024
	NN-dPET	4.586 ± 1.077	0.763 ± 0.024	0.753 ± 0.024
	p-metcorIDIF	4.423 ± 0.633	0.685 ± 0.034	0.675 ± 0.034

Supplementary Table 9. Comparison of parameter estimates for Dataset 2. Values are median ± Median Absolute Deviation (MAD). Inter-subject correlation (ρ) and concordance (CCC) metrics are reported.Abbreviations: 1T1K implemented as in[14]

Parameter	Method	Median ± MAD	Inter-subject ρ	Inter-subject CCC
K_1	True	0.096 ± 0.020	–	–
	NN-unBlood	0.095 ± 0.016	0.946 ± 0.006	0.942 ± 0.007
	p-metcorBlood	0.081 ± 0.014	0.970 ± 0.004	0.690 ± 0.013
	NN-dPET	0.099 ± 0.021	0.534 ± 0.042	0.522 ± 0.041
	p-metcorIDIF	0.478 ± 0.079	0.655 ± 0.030	0.026 ± 0.003
K_1-1T1K	True	0.093 ± 0.015	–	–
	NN-unBlood	0.092 ± 0.013	0.982 ± 0.001	0.979 ± 0.001
	p-metcorBlood	0.077 ± 0.012	0.983 ± 0.001	0.703 ± 0.010
	NN-dPET	0.098 ± 0.014	0.587 ± 0.029	0.579 ± 0.029
	p-metcorIDIF	0.298 ± 0.037	0.617 ± 0.030	0.033 ± 0.003
k_2	True	0.107 ± 0.013	–	–
	NN-unBlood	0.091 ± 0.026	0.420 ± 0.085	0.378 ± 0.073
	p-metcorBlood	0.098 ± 0.014	0.749 ± 0.062	0.711 ± 0.061
	NN-dPET	0.083 ± 0.022	0.001 ± 0.072	0.001 ± 0.050
	p-metcorIDIF	0.348 ± 0.066	0.297 ± 0.040	0.031 ± 0.006
k_3	True	0.065 ± 0.018	–	–
	NN-unBlood	0.065 ± 0.027	0.336 ± 0.072	0.306 ± 0.064
	p-metcorBlood	0.069 ± 0.015	0.578 ± 0.089	0.563 ± 0.084
	NN-dPET	0.054 ± 0.029	0.157 ± 0.095	0.123 ± 0.070
	p-metcorIDIF	0.065 ± 0.009	0.218 ± 0.074	0.126 ± 0.043
k_4	True	0.027 ± 0.007	–	–
	NN-unBlood	0.031 ± 0.009	0.265 ± 0.067	0.062 ± 0.024
	p-metcorBlood	0.026 ± 0.005	0.533 ± 0.053	0.489 ± 0.051
	NN-dPET	0.026 ± 0.007	0.253 ± 0.057	0.123 ± 0.047
	p-metcorIDIF	0.021 ± 0.002	0.426 ± 0.048	0.166 ± 0.020
V_b	True	0.035 ± 0.006	–	–
	NN-unBlood	0.035 ± 0.006	0.972 ± 0.005	0.967 ± 0.006
	p-metcorBlood	0.036 ± 0.006	0.991 ± 0.002	0.989 ± 0.002
	NN-dPET	0.038 ± 0.006	0.364 ± 0.065	0.333 ± 0.057
	p-metcorIDIF	0.080 ± 0.016	0.097 ± 0.059	0.017 ± 0.010
V_T	True	3.090 ± 0.541	–	–
	NN-unBlood	3.180 ± 0.624	0.734 ± 0.084	0.717 ± 0.097
	p-metcorBlood	2.928 ± 0.719	0.846 ± 0.012	0.830 ± 0.013
	NN-dPET	3.419 ± 0.679	0.498 ± 0.127	0.474 ± 0.142
	p-metcorIDIF	5.505 ± 0.891	0.568 ± 0.038	0.179 ± 0.012

Supplementary Table 10. Comparison of PET metrics between PT (KOA) and HC groups for the main study and Dataset 3. Values are mean (μ), standard deviation (σ), and coefficient of variation (cv). Group differences are shown with p-values, and effect sizes are reported as Cohen's d with lower (LCI) and upper (UCI) 95% confidence intervals. Legend: n.s. = non significant ($p > 0.05$); positive d indicates higher values in PT relative to HC.

	Cortex	WM	Thalamus	Pituitary	ACC	PCC	S1
SUVR (main study)	($n_{\mathrm{PT}} = 41$, $n_{\mathrm{HC}} = 22$)						
μ_{PT}	1.48	1.48	1.81	1.53	1.66	1.68	1.35
μ_{HC}	1.31	1.30	1.60	1.29	1.47	1.48	1.20
σ_{PT}	0.27	0.24	0.28	0.30	0.29	0.31	0.28
σ_{HC}	0.18	0.13	0.19	0.22	0.18	0.22	0.18
cv (σ/μ%)	17.8	15.6	15.7	20.5	16.9	18.2	19.8
p	< 0.001	< 0.001	< 0.001	< 0.01	< 0.001	< 0.001	0.001
d	0.84	0.89	0.92	0.63	0.87	0.93	0.72
95% LCI	0.29	0.33	0.37	0.09	0.32	0.37	0.17
95% UCI	1.39	1.44	1.48	1.17	1.42	1.48	1.26
SUVR (Dataset 3)	($n_{\mathrm{PT}} = 21$, $n_{\mathrm{HC}} = 11$)						
μ_{PT}	1.48	1.48	1.82	1.49	1.67	1.70	1.37
μ_{HC}	1.30	1.28	1.60	1.35	1.45	1.47	1.18
σ_{PT}	0.27	0.22	0.28	0.28	0.29	0.31	0.29
σ_{HC}	0.19	0.14	0.18	0.25	0.19	0.22	0.17
cv (σ/μ%)	18.1	15.5	15.3	19.1	17.6	18.3	20.6
p	0.011	0.010	0.002	0.077	0.004	0.005	0.014
d	0.79	0.80	0.96	0.55	0.90	0.87	0.76
95% LCI	0.01	0.01	0.16	-0.23	0.10	0.07	-0.03
95% UCI	1.58	1.59	1.76	1.32	1.69	1.66	1.55
DVR (Dataset 3)	($n_{\mathrm{PT}} = 21$, $n_{\mathrm{HC}} = 11$)						
μ_{PT}	1.44	1.36	1.72	1.98	1.56	1.58	1.33
μ_{HC}	1.25	1.16	1.49	1.84	1.34	1.34	1.14
σ_{PT}	0.18	0.14	0.21	0.39	0.23	0.20	0.20
σ_{HC}	0.19	0.14	0.21	0.42	0.18	0.20	0.16
cv (σ/μ%)	14.6	13.1	14.3	20.6	16.0	15.3	16.3
p	< 0.001	< 0.001	0.001	0.115	< 0.001	< 0.001	< 0.001
d	1.08	1.08	1.02	0.49	1.07	1.11	1.04
95% LCI	0.27	0.27	0.21	-0.28	0.26	0.29	0.23
95% UCI	1.90	1.90	1.82	1.26	1.88	1.92	1.85
V_T (Dataset 3)	($n_{\mathrm{PT}} = 21$, $n_{\mathrm{HC}} = 11$)						
μ_{PT}	3.13	2.96	3.76	4.18	3.41	3.45	2.89
μ_{HC}	3.26	3.05	3.93	4.91	3.53	3.53	2.96
σ_{PT}	1.09	1.04	1.39	1.30	1.31	1.30	1.00
σ_{HC}	0.83	0.79	1.14	2.03	1.00	1.00	0.74
cv (σ/μ%)	31.4	31.8	33.9	36.0	34.7	34.1	31.3
p	0.73	0.88	0.72	0.88	0.70	0.74	0.53
d	0.11	0.05	0.11	0.05	0.12	0.10	0.19
95% LCI	-0.65	-0.71	-0.65	-0.71	-0.64	-0.66	-0.57
95% UCI	0.87	0.81	0.87	0.81	0.88	0.86	0.95

Supplementary Pseudo-code

Algorithm 1: ASTRA Inference and Training Pipeline: NN-dPET

Input: Dynamic PET volume $X \in \mathbb{R}^{T \times H \times W \times D}$; Clinical covariates c
Output: Metabolite-corrected plasma curve $\hat{y}$; predictive uncertainty $\hat{\sigma}$

Module 1 (M1) – Dimensionality Reduction;
Split X into fixed-size 3D patches across time;
Encode patches with Conv3D layers + GroupNorm;
Inject FiLM conditioning from c into each encoder block;
Aggregate latent sequence to form temporally ordered embedding z_{M1};

Module 2 (M2) – Signal Extraction Network;
Apply 3D temporal convolutions to z_{M1};
Process through transformer encoder blocks ;
Inject FiLM conditioning at each transformer block;
Project outputs via Dense & Projection layers to predict uncorrected plasma curve $\hat{p}$;

Module 3 (M3) – Metabolite Correction Network;
Project $\hat{p}$ to transformer model dimension via fully connected layers;
Add temporal position encodings to the projected sequence;
Process through transformer encoder blocks;
Inject FiLM conditioning at each transformer block;
Project via a 2-layer MLP to predict mean $\mu(t)$ and standard deviation $\sigma(t)$;

Output;
Return predicted metabolite-corrected curve $\hat{y}$ and associated uncertainty $\hat{\sigma}$;

Algorithm 2: ASTRA Inference and Training Pipeline: NN-unBlood

Input: Uncorrected plasma curve $p \in \mathbb{R}^{T}$; Clinical covariates c
Output: Metabolite-corrected plasma curve $\hat{y}$; predictive uncertainty $\hat{\sigma}$

Module 3 (M3) – Metabolite Correction Network;
Project p to transformer model dimension via fully connected layers;
Add temporal position encodings to the projected sequence;
Process through transformer encoder blocks;
Inject FiLM conditioning at each transformer block;
Project via a 2-layer MLP to predict mean $\mu(t)$ and standard deviation $\sigma(t)$;

Output;
Return predicted metabolite-corrected curve $\hat{y}$ and associated uncertainty $\hat{\sigma}$;