International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
SJIF (2025): 8.036

Organ-Age: A Multimodal Fusion of Transcriptomic and Radiological Signals for Organ-Resolved Biological Age Estimation

Arnav Sangle*, Independent Researcher, Ranchview High School, Irving TX, USA
arnavsangle08@gmail.com
*Corresponding Author

Abstract

This study proposes Organ-Age, a multimodal framework for estimating organ-resolved biological age by integrating transcriptomic and radiological data. Using bulk RNA-seq (GTEx), chest radiographs (CheXpert), and brain MRI (IXI), modality-specific encoders generate embeddings aligned through contrastive learning and fused via a transformer-based architecture. The model outputs probabilistic organ-level age estimates, evaluated on a combined cohort exceeding 190,000 samples. The aligned multimodal approach achieved a mean absolute error of approximately 9.3 years, outperforming unimodal and unaligned fusion baselines. Residual differences between predicted and chronological age revealed heterogeneous aging patterns across organs. While results suggest potential for localized aging analysis, interpretability and causal inference remain limited by dataset heterogeneity and absence of matched multimodal samples. These findings support further exploration of multimodal representation learning for organ-specific aging assessment.

Keywords: biological age, organ aging, multimodal learning, contrastive learning, contrastive alignment, transcriptomics, fusion transformer, interpretability, biological aging biomarkers, probabilistic regression, multimodal representation learning

1 Introduction

Biological aging involves progressive functional decline at the molecular, cellular, and organ levels. Chronological age is a crude proxy for this process: individuals of the same calendar age vary widely in healthspan, disease burden, and rate of physiological deterioration. Recognizing this, a growing body of work has sought to estimate biological age, a quantity that reflects physiological state rather than time since birth [1, 2].

The most established biological age predictors are epigenetic clocks, which use DNA methylation patterns to estimate a single, organism-wide age value. These clocks correlate with mortality, chronic disease risk, and functional decline [3]. They have been instrumental in showing that aging is, in principle, quantifiable from molecular data. However, because they collapse aging into one number, they say little about how individual tissues or organs age relative to one another—a distinction that matters clinically, since age-related pathology is often organ-specific.

A growing literature confirms that aging is not synchronous across the body. Lungs in chronic smokers and livers in heavy drinkers can deteriorate well ahead of other organs in the same individual. More broadly, tissue-resolved profiling studies document wide variation in the pace and character of age-associated molecular change across organ systems [4]. Recent large-scale studies reinforce this view: Tian et al. showed that heterogeneous aging across multiple organ systems predicts chronic disease and mortality in the UK Biobank [5], while Oh et al. demonstrated that plasma proteomic signatures can resolve organ-specific aging trajectories linked to disease risk [6]. Despite this growing evidence, few computational approaches attempt to estimate organ-specific biological age directly. Most existing models work from a single data type—gene expression, imaging, or methylation—and therefore capture only part of the picture.

On the imaging side, deep learning models trained on chest radiographs and brain MRI show that macrostructural aging features—changes in skeletal morphology, soft-tissue composition, and brain volume—can be inferred from medical images with strong predictive accuracy [7, 8]. These structural cues are largely invisible to molecular

clocks, and vice versa. Yet imaging-based age predictors are typically developed in isolation, with no explicit connection to underlying molecular state.

Multimodal learning offers a natural way to bridge this gap. Contrastive objectives such as InfoNCE can pull embeddings from different data types into a common representation space, preserving structure within each modality while encouraging cross-modal agreement [9–11]. Transformer-based architectures add the ability to model interactions across modalities through attention [12], and recent systems like Perceiver IO and FLAVA demonstrate that diverse inputs can be merged through a shared latent bottleneck without requiring pixel-level or token-level alignment [13, 14]. Wang et al. recently applied a multimodal image Transformer combining facial, fundus, and tongue photographs to estimate biological age and predict chronic disease, achieving strong correlation with chronological age across independent cohorts [15]. Separately, probabilistic regression emphasizes the value of predicting uncertainty alongside point estimates, which is relevant for targets like biological age that are inherently noisy and never directly observed [16]. Taken together, these lines of work suggest that biological age may be better understood as a composite quantity shaped by the interplay of molecular and structural processes than as something derivable from any single measurement.

With this motivation, this study presents Organ-Age, a framework for estimating organ-level biological age by jointly modeling transcriptomic and radiological signals. Separate encoders produce embeddings from RNA-seq data, chest radiographs, and structural MRI scans; these embeddings are pulled into a shared latent space through contrastive alignment and then merged by a transformer-based fusion module to yield organ-specific age predictions. A central output of the work is the residual deviation between predicted organ age and chronological age, which can flag accelerated or decelerated aging in specific tissues.

Figure 1 outlines the Organ-Age pipeline. The framework assumes that organ-specific biological age reflects the joint contribution of molecular and structural signals and that aligning representations from these domains can surface aging patterns invisible to any single modality. Beyond population-level evaluation, Organ-Age also supports per-subject interpretation through individualized organ-age panels and gene-level attribution summaries, which trace organ-age deviations back to candidate transcriptomic drivers. Specifically, this work aims to: (1) demonstrate that contrastive alignment of transcriptomic and radiological embeddings improves organ-level age prediction relative to unimodal and unaligned baselines; (2) quantify organ-specific aging heterogeneity through calibrated residual analysis; and (3) provide interpretable per-subject organ-age profiles with uncertainty estimates and gene-level attribution.

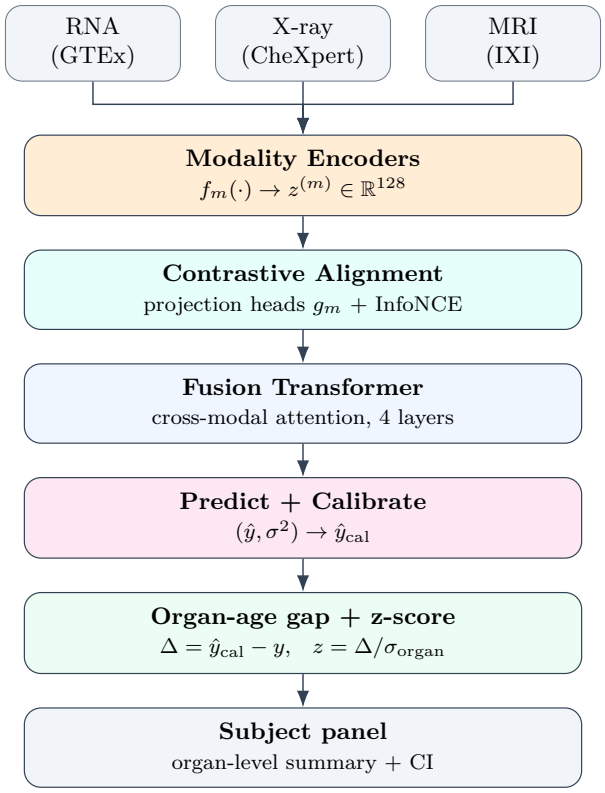


Figure 1: Organ-Age pipeline. Three modality-specific encoders produce embeddings that are aligned via contrastive learning, fused by a transformer, and post-processed into calibrated organ-age predictions with uncertainty intervals and per-subject summaries.

2 Methodology

Three publicly available datasets supply the molecular and structural aging signals used in this work. Transcriptomic data come from the Genotype-Tissue Expression (GTEx) v10 release, which contains bulk RNA-seq profiles across dozens of human tissues sampled from adult donors [17]. Radiological data come from CheXpert, a large collection of frontal chest radiographs with associated demographic metadata [18]. Structural brain imaging comes from the IXI dataset, which provides T1-weighted MRI scans from healthy adults spanning a broad age range [19]. Samples without reliable age labels or that failed quality-control filters were dropped before any modeling.

Each data type was preprocessed according to standard domain-specific protocols. Gene expression counts were variance-stabilized following the DESeq2 approach [20], and batch effects from site and technical sources were cor-

rected with ComBat [21]. Chest radiographs were resized and intensity-normalized, then passed through convolutional backbones initialized from pretrained ResNet weights [22]; transfer-learning practices from BiT [23] guided the fine-tuning strategy. MRI volumes were skull-stripped, intensity-normalized, and nonlinearly registered to a common template with ANTs [24], supplemented by standard FSL preprocessing steps [25].

A separate encoder network handles each modality. For transcriptomics, a feedforward network maps normalized expression vectors to a fixed-length embedding. For chest X-rays and MRI scans, convolutional networks—adapted to 2-D and 3-D inputs, respectively—serve the same purpose. All three encoders output embeddings of the same dimensionality so that they can be compared and merged downstream.

To bring modality-specific embeddings into a common space, learned projection heads were appended to each encoder and trained with an InfoNCE-style contrastive objective [9], borrowing design choices from SimCLR and CLIP [10, 11]. Pairwise similarity is measured by cosine similarity, which depends on the angle between vectors and is insensitive to their magnitude:

$$\mathrm{sim}(u, v) = \frac{u^\top v}{\|u\|\|v\|} \tag{1}$$

Alignment was optimized by minimizing the contrastive loss

$$\mathcal{L}_{\mathrm{InfoNCE}} = -\log \frac{\exp(\mathrm{sim}(u, u')/\tau)}{\sum_j \exp(\mathrm{sim}(u, u_j)/\tau)} \tag{2}$$

which pulls embeddings from the same biological context closer together while pushing apart those from unrelated samples. The purpose is not to erase modality-specific information but to reduce modality identity as the dominant axis of variation, so that age-related geometry can be fused reliably.

Once aligned, modality embeddings are combined by a transformer-based fusion module. Instead of simple concatenation or prediction averaging, the fusion model uses attention to learn cross-modal interactions, letting molecular and radiological information contribute in a data-dependent way. Given aligned embeddings $\{z^{(1)}, z^{(2)}, \ldots, z^{(M)}\}$ from M modalities, the fused representation is

$$z_{\mathrm{fusion}} = f_{\mathrm{Transformer}}\left(z^{(1)}, z^{(2)}, \ldots, z^{(M)}\right) \tag{3}$$

The design follows Perceiver IO and FLAVA [13, 14], adapted to operate on continuous biomedical embeddings.

The fused representation is fed to a probabilistic regression head that outputs both a point estimate and a learned uncertainty term, rather than a bare scalar prediction. Concretely, the head produces

$$(\mu, \sigma^2) = g_\theta(z_{\mathrm{fusion}}) \tag{4}$$

where $g_\theta(z_{\mathrm{fusion}})$ is the regression head with parameters θ. The training objective is a Gaussian negative log-likelihood [16]

$$\mathcal{L}_{\mathrm{NLL}} = \frac{1}{2}\log\sigma^2 + \frac{(\mu - y)^2}{2\sigma^2} \tag{5}$$

so that the model is incentivized to widen its uncertainty estimate when its predictions are poor, rather than absorbing all error into the mean.

Training follows a staged curriculum: first the modality-specific encoders are trained independently, then the contrastive alignment and fusion stages are optimized jointly. The combined loss is

$$\mathcal{L}_{\mathrm{total}} = \lambda_{\mathrm{align}}\mathcal{L}_{\mathrm{InfoNCE}} + \lambda_{\mathrm{reg}}\mathcal{L}_{\mathrm{NLL}} \tag{6}$$

where λ_{align} and λ_{reg} weight the two terms. Both coefficients were set on a held-out validation split and kept fixed thereafter.

Regularization during optimization draws on the variational information bottleneck [26]. The full training procedure—encoder pretraining, contrastive alignment, and fusion—was run sequentially, with each stage warm-starting from the previous checkpoint.

Beyond population-level metrics, Organ-Age provides per-subject outputs: individualized organ-deviation panels and gene-level attribution summaries that trace organ-age residuals back to specific transcripts. Statistical significance of per-organ age-gap distributions was assessed using Wilcoxon signed-rank tests, and bootstrap resampling (10,000 replicates) was used to construct 95% confidence intervals on MAE estimates.

3 Results and Discussion

The model was evaluated in three configurations—unimodal, naïve multimodal fusion, and contrastively aligned multimodal fusion—to isolate what each component contributes. Performance is reported as mean absolute error (MAE) and mean squared error (MSE). All metrics are computed on held-out test splits; training and validation data were used exclusively for model selection and hyperparameter tuning.

3.1 Model performance

Each modality was first examined in isolation. The RNA encoder, trained on GTEx expression data ($n = 7{,}378$), achieved an MAE of 9.62 years (MSE = 156.4). Prediction variance increased at older ages, consistent with the documented rise in transcriptomic heterogeneity among

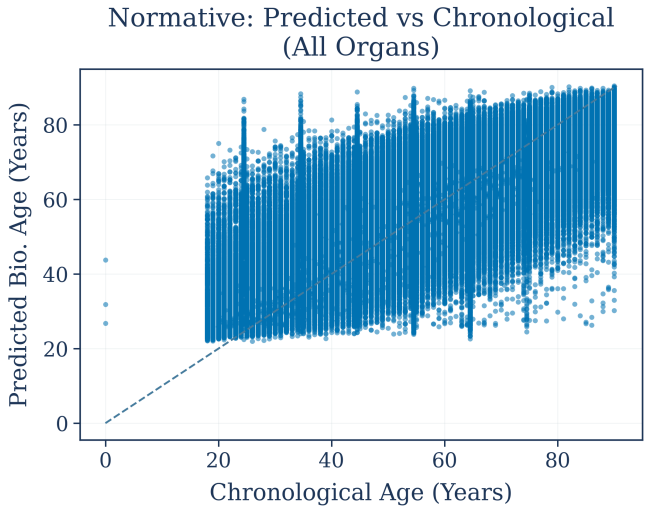


Figure 2: Overall prediction behavior across the population (normative setting). Predicted biological age versus chronological age for the aggregated cohort.

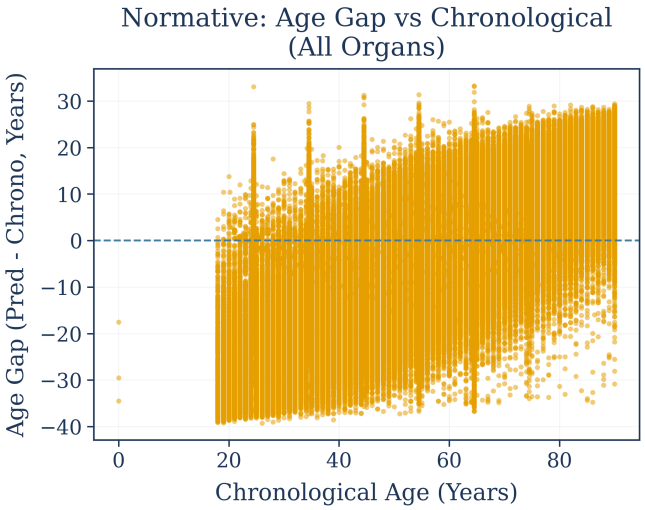


Figure 3: Age-gap behavior across the population (normative setting). Difference between predicted biological age and chronological age as a function of chronological age.

older individuals. The encoder captured molecular aging correlates across tissues but lacked structural context, limiting its ability to resolve anatomical features of organ aging.

The chest X-ray encoder (n = 187,825) achieved an MAE of 10.92 years (MSE = 186.2), learning macrostructural aging cues including skeletal morphology, cardiac silhouette shape, and soft-tissue composition changes. However, predicted ages clustered within a compressed range around the population mean (effective prediction range ∼25–65 years versus a true span of 18–90), indicating that the model captured a robust population trend while attenuating individual-level variation. The MRI encoder (n = 563) showed the highest MAE among unimodal baselines at 14.93 years (MSE = 312.2), reflecting the challenge of generalizing structural brain signals to cross-tissue age estimation without a shared latent space; as shown below, alignment substantially improves MRI performance.

As a first multimodal baseline, a naïve fusion model was constructed that concatenates modality-specific embeddings without prior alignment. Across the full dataset this achieved an MAE of 10.88 years (MSE = 185.4)—comparable to the X-ray unimodal result and worse than RNA alone, confirming that naïve concatenation does not reliably leverage cross-modal complementarity without prior alignment. However, UMAP projections of the concatenated embeddings showed that points still clustered primarily by modality rather than by age (Section 3.2), reflecting unstable fusion behavior.

Introducing contrastive alignment before the fusion stage reduced inter-modality variance in the latent space

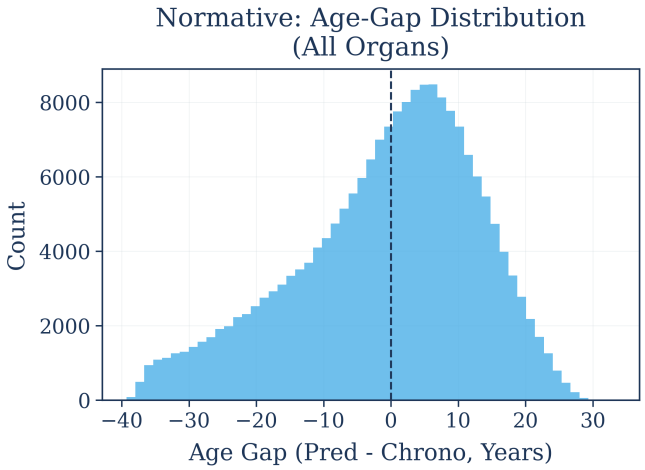


Figure 4: Organ-age gap distribution (normative setting). Histogram of $\Delta = \hat{y} - \text{Age}_{\text{true}}$ summarizing the distribution of accelerated (positive) and decelerated (negative) aging deviations across the cohort.

and yielded consistent improvements across all metrics. The aligned fusion model achieved an MAE of 9.30 years (MSE = 138.0) across the full 195,766-sample dataset. Table 1 summarizes the ablation: aligned fusion reduced MAE by 14.6% relative to naïve fusion. The MRI modality showed the largest per-modality gain, with MAE dropping from 14.93 to 6.21 years (−58%) after alignment, compared with 14.7% for X-ray and 4.8% for RNA.

The improvement held across all three modalities when evaluated per-modality within the aligned framework

Table 1: Ablation: unimodal, naïve fusion, and aligned fusion. All metrics are on held-out test splits. Unimodal results use per-modality evaluation of the naïve fusion model on subjects with access to only that input type. ΔMAE reports the relative MAE reduction from naïve fusion to aligned fusion.

Configuration	MSE	MAE (yr)
RNA only	156.4	9.62
X-ray only	186.2	10.92
MRI only	312.2	14.93
Naïve fusion	185.4	10.88
Aligned fusion	138.0	9.30

Table 2: Performance of contrastively aligned multimodal fusion, by modality. MSE and MAE on held-out test splits. Per-modality results confirm that alignment benefits all input types.

Modality	N	MSE	MAE (yr)
All	195,766	138.00	9.30
RNA	7,378	139.44	9.16
X-Ray	187,825	138.18	9.31
MRI	563	61.92	6.21

(Table 2), indicating that alignment makes each modality a more consistent contributor to the fused representation rather than one modality dominating. The MRI subset achieved the lowest MAE (6.21 years), likely reflecting both the higher intrinsic signal-to-noise of structural brain imaging and the relative homogeneity of the IXI cohort.

To visualize what alignment does to the latent space, unaligned and aligned embeddings were projected with UMAP [27]. Before alignment, the first two UMAP components separated points by modality, with modality identity explaining the majority of embedding variance. After alignment, embeddings organized along age-related gradients rather than by modality identity, confirming that the contrastive objective shifted the dominant axis of variation from domain identity to aging structure.

The primary derived metric is the organ-age gap:

$$\Delta = \hat{y} - \text{Age}_{\text{true}} \tag{7}$$

In the aligned model, predicted ages tracked chronological age in a tighter band (residual SD = 11.8 years, versus 13.6 in the naïve fusion model), a reduction driven primarily by alignment's stabilizing effect on the MRI modality. Age-bin analysis reveals the expected regression-to-mean pattern: prediction errors are highest for the youngest subjects (20–40 age range: MSE = 300) and lowest for the 60–80 group (MSE = 85), consistent

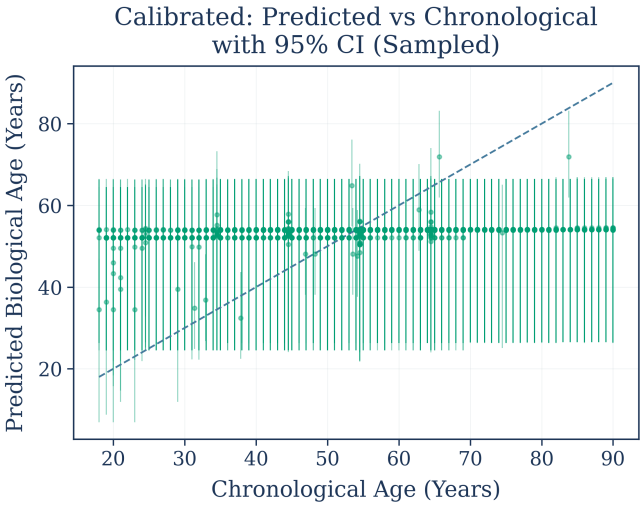


Figure 5: Calibrated organ-age prediction with uncertainty. Calibrated predicted biological age with a 95% confidence interval compared to chronological age.

with cohort-trained age predictors attenuating extreme values toward the population mean. This matters because organ-resolved interpretation depends on stable residual structure; if residuals drift systematically with age or modality availability, organ-age deltas become difficult to compare across individuals.

3.2 Calibration and uncertainty

Post-hoc calibration was applied to put predicted ages on a scale more directly comparable to chronological age and to make residuals comparable across organs. Figure 5 shows the result. After calibration, a systematic regression-to-mean bias is evident: predicted biological ages are over-estimated for young subjects (+27 years in the 20–30 bin, +18 years in the 30–40 bin) and under-estimated for older subjects (−10 years in the 60–70 bin, −20 years in the 70–80 bin). This reflects the well-known tendency of cohort-trained age predictors to compress predictions toward the population mean. The 95% confidence intervals are approximately ±20 years across all age bins, providing context for organ-age deviations and motivating the organ-resolved z-score framing, which accounts for this age-dependent structure.

To quantify that organ aging is not uniform, calibrated results are broken out for four representative organs. Table 3 reports per-organ MAE, residual standard deviation, and mean age-gap ($\bar{\Delta}$).

Brain and brain cortex show the tightest residual distributions (SD = 6.48 and 10.14 years, respectively), consistent with the lower measurement noise and more constrained biological variability of structural neuroimaging. Lung shows the widest dispersion (SD = 17.29 years),

Table 3: Per-organ prediction metrics (calibrated). MAE, residual standard deviation (Res. SD), and mean age gap ($\bar{\Delta}$) for four representative organs. Differences in dispersion and bias across organs confirm that aging is not uniform.

Organ	N	MAE (yr)	Res. SD (yr)	$\bar{\Delta}$ (yr)
Brain	563	5.01	6.48	0.0
Brain cortex	270	7.95	10.14	0.0
Heart	913	8.19	10.85	0.0
Lung	188,429	15.50	17.29	−6.93

reflecting both greater biological heterogeneity—driven in part by exposure variation such as smoking—and the noisier transcriptomic signal in this tissue. Notably, lung carries a substantial negative mean age gap ($\bar{\Delta} = -6.93$ years), indicating that predicted lung ages systematically under-estimate chronological age in this cohort.

To assess statistical significance, Wilcoxon signed-rank tests were applied to each organ's calibrated age-gap distribution under the null hypothesis that the median organ-age delta equals zero. Lung ($p < 0.001$) and brain cortex ($p = 5.1 \times 10^{-5}$) showed statistically significant deviations from zero, while brain ($p = 0.76$) and heart ($p = 0.82$) did not, confirming that the observed lung under-estimation is a systematic effect rather than sampling noise. Bootstrap 95% confidence intervals (10,000 replicates) on per-organ MAE further confirm that inter-organ differences are robust: brain MAE = 5.01 [4.68, 5.35], heart MAE = 8.19 [7.73, 8.65], and lung MAE = 15.50 [15.45, 15.54] years. Pairwise bootstrap comparisons of MAE between organs were all statistically significant ($p < 0.05$), indicating that prediction accuracy genuinely varies across organ systems rather than reflecting chance differences.

Figure 6 confirms these patterns visually. All four organs exhibit roughly monotonic predicted-versus-chronological-age relationships, but they differ in slope attenuation (brain predictions span ∼69% of the true age range; lung predictions are heavily compressed), residual structure, and age-gap density shape. These organ-specific patterns would be invisible in a pooled analysis and motivate the individualized interpretation that follows.

3.3 Individual-level interpretation

Organ-age deltas should be interpreted as correlational signals, not diagnoses. Even so, the organ-resolved structure makes it possible to separate localized acceleration from global drift, which is relevant for exposures that disproportionately affect specific organs.

For lungs, chronic smoking is a canonical example. Smoking alters airway and parenchymal structure over

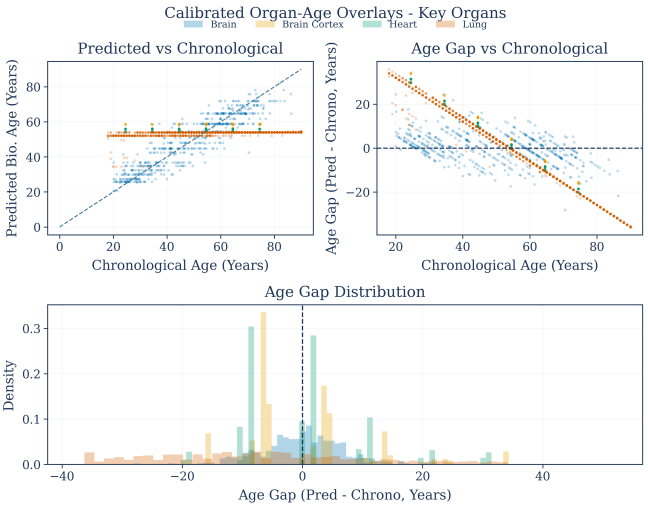


Figure 6: Calibrated organ-age overlays for key organs. Compact three-view summary for four representative organs (brain, brain cortex, heart, lung): *top-left* overlay of predicted biological age vs. chronological age, *top-right* overlay of age-gap trajectories $\Delta = \hat{y} - \text{Age}_{\text{true}}$, and *bottom* density-normalized age-gap distributions.

years, producing radiographic patterns and transcriptomic changes tied to inflammation, extracellular matrix remodeling, and stress-response programs. In a fused representation, these molecular and structural cues can reinforce one another. An older-appearing lung profile can thus emerge even when organism-wide age remains near expected, because the signal is localized rather than systemic.

For liver, chronic alcohol burden and metabolic strain are analogous. Alcohol exposure and metabolic dysregulation drive transcriptomic shifts in lipid handling, oxidative stress, and inflammatory pathways, alongside structural consequences that can manifest as altered tissue composition and downstream systemic effects. The key point is not causal attribution, but that multi-signal fusion supports a *localized deviation* framework: the model can register a coherent liver-specific age gap even when other organs remain nominal.

This interpretation has a practical implication: organ-resolved age gaps provide a readable, organ-local summary that can be used to prioritize follow-up. The output is not a clinical label. It is a structured flag: *this organ deviates from age-matched expectation by* Δ *years with confidence interval* $[\Delta - z_\alpha\sigma,\ \Delta + z_\alpha\sigma]$.

While cohort-level metrics validate overall model behavior, individualized interpretation is the intended endpoint of Organ-Age. A single representative case study illustrates how calibrated organ-age outputs translate into organ deltas, uncertainty intervals, and standardized

z-scores.

Subject GTEX-1117F (chronological age ≈ 64.5 years) was selected as a representative case because tissue was available for multiple organs, enabling a cross-organ comparison within a single individual. Figure 7(a) shows the per-organ deviation profile. Skeletal muscle exhibited the largest negative deviation ($\Delta = -14.1$ years, $z = -1.09$), while brain cortex showed a more moderate offset ($\Delta = -5.7$ years, $z = -0.56$). All sampled organs for this subject fell below chronological age, yielding a mean organ-age gap of $\bar{\Delta} = -10.5$ years across the seven profiled tissues—consistent with a globally younger-appearing molecular and structural profile.

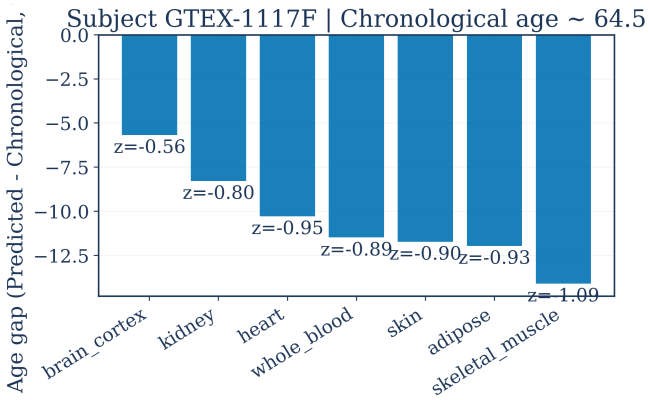
The radar profile in Figure 7(b) visualizes the z-score pattern across all profiled organs, confirming that deviations are globally negative rather than organ-specific. Calibrated 95% confidence intervals (not shown) add further nuance: the skeletal muscle deviation of −14.1 years carries a wide CI (∼22–65 years), warranting caution, whereas the kidney deviation of −8.3 years has a narrower interval (∼35–72 years), indicating a more stable estimate. All organs for this subject fall below chronological age, supporting interpretation as a systemic rather than organ-local phenomenon.

3.4 Molecular interpretability

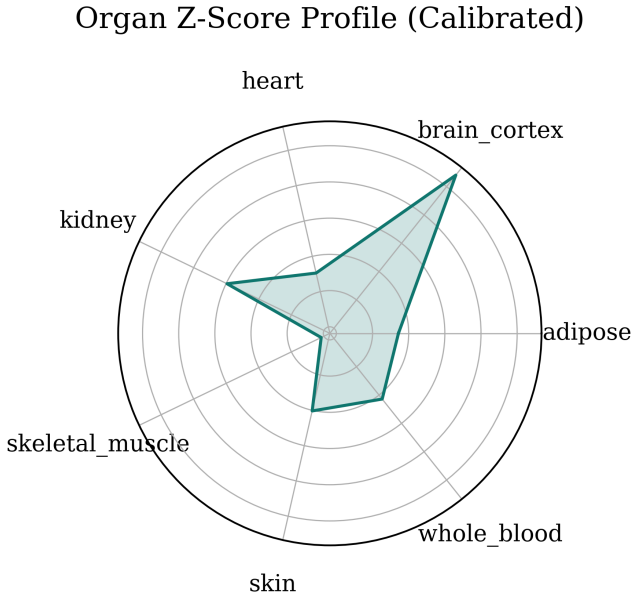
A model that predicts organ age accurately but offers no molecular insight is of limited scientific value. The organ-age panels show *which* organs deviate; the attribution extension asks *why*, by identifying which genes and latent factors drive the predictions. Two complementary analyses are reported below.

First, gene-ranking analyses list individual genes whose expression patterns are most strongly associated with predicted organ age, yielding a direct feature-level summary of candidate contributors. Second, integrated gradients computed over the latent RNA dimensions show how the model compresses transcriptomic information internally and which latent factors matter most for each organ. Between them, these views move from black-box prediction toward concrete molecular hypotheses.

Figure 8 shows that attribution is distributed broadly across the top 20 genes for all profiled organs. The top 5 genes account for ∼28% of total attribution mass among the top 20 for liver, kidney, brain cortex, and heart alike, indicating that no single gene dominates and that the model integrates signal from many transcriptomic contributors. Five genes (gene_202, gene_394, gene_891, gene_919, gene_931) appear in the top 20 for all four organs, suggesting a shared core of aging-correlated transcripts, while organ-specific genes provide additional



(a) Per-organ age-gap bars with z-score annotations.



(b) Polar radar chart of calibrated organ z-scores.

Figure 7: Representative case study: individualized organ-age profile for GTEX-1117F. (a) Per-organ deviation from chronological age with calibrated z-scores. (b) Radar profile showing the pattern of organ-level z-scores; all organs fall below zero, indicating globally younger biological age.

tissue-contextualized signal.

These rankings should be read as associations with the model output rather than causal drivers of aging. Even so, they provide a practical starting point for follow-up: pathway enrichment, comparison to published aging gene sets, and targeted validation in longitudinal or disease-specific cohorts.

Figure 9 shows that attribution is distributed broadly across latent RNA dimensions rather than concentrated in a small subset. For liver, the top 3 latent dimensions capture 3.6% of total integrated-gradient magnitude; for kidney, 4.2%; for brain cortex, 4.1%. This near-

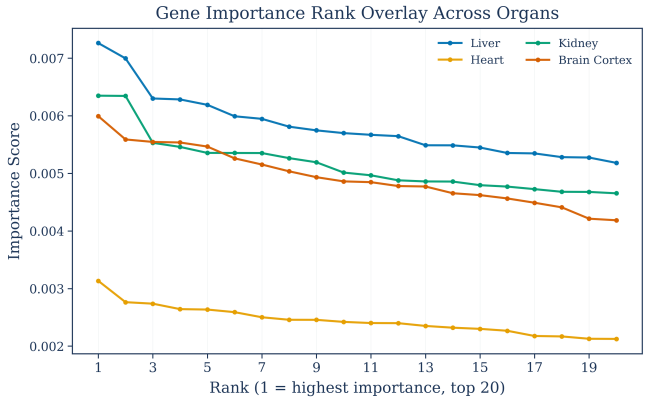


Figure 8: Gene-level attribution: rank overlay of top-gene importance by organ. Importance scores across ranks (top 20 genes per organ) are overlaid for representative organs, enabling compact cross-organ comparison of attribution concentration and tail decay.

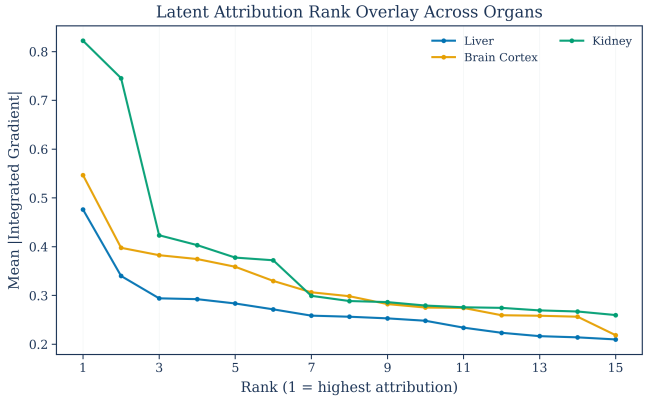


Figure 9: Latent-space feature importance (RNA): cross-organ rank overlay. Integrated-gradient-derived latent importance over rank for representative organs, highlighting how attribution mass concentrates in a subset of dimensions.

uniform distribution across the 256-dimensional latent space suggests the model spreads transcriptomic aging information broadly rather than compressing it into a compact bottleneck. Latent dimension 114 appears in the top 5 for liver, kidney, and brain cortex, representing one shared contributor, while organ-preferential dimensions include 211 and 204 for liver; 144, 215, and 224 for kidney; and 92, 126, and 132 for brain cortex, consistent with a decomposition into weakly shared systemic factors and organ-specific programs.

Figure 10 zooms in on liver, where the cumulative attribution curve rises gradually: the top 5 dimensions capture 5.8% of total integrated-gradient mass, and the top 10 reach 10.7%. This diffuse pattern indicates that liver age predictions draw on a broad transcriptomic

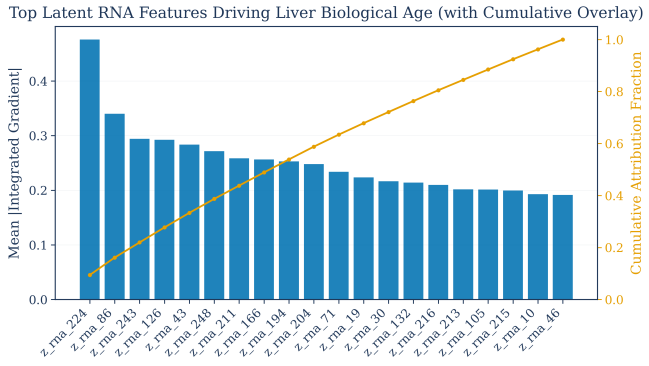


Figure 10: Latent RNA features driving liver biological age with cumulative overlay. Top latent RNA dimensions (bars) and cumulative attribution fraction (line) for liver age prediction.

signal distributed across the latent space rather than a compact bottleneck. Each set of top dimensions nonetheless offers an entry point for deeper investigation through gene loadings, pathway correlations, and comparison to known hepatic aging programs.

Taken together, the organ-age panels and gene-level attributions connect organ-age deviations to plausible molecular contributors, moving the framework from prediction toward hypothesis generation.

3.5 Discussion

The experiments support a clear conclusion: combining molecular and radiological data in a shared representation improves organ-level biological age estimation. The key ingredient is contrastive alignment. Without alignment, embeddings occupy distinct regions of latent space, and the fusion module integrates them inconsistently. With alignment, transcriptomic, chest-X-ray, and MRI features intermix and organize along age gradients, which makes fusion behavior more stable. This echoes a broader result in multimodal learning: representation alignment is often a prerequisite for reliable cross-modal integration [28].

Quantitatively, aligned fusion reduced MAE by 14.6% over naïve fusion (Table 1). The gain was not driven by a single modality: per-modality evaluation within the aligned framework showed consistent improvements across RNA (4.8%), X-ray (14.7%), and MRI (58%) inputs (Table 2), indicating that alignment makes each data type a better contributor to the joint representation. The residual standard deviation dropped from 13.6 years (naïve fusion) to 11.8 years (aligned fusion). These improvements in residual stability are prerequisite for reliable organ-resolved interpretation, because organ-age deltas are only meaningful if the residual structure is consistent across the age range.

Relative to existing biological age predictors, Organ-Age adds organ-level resolution. Epigenetic and transcriptomic clocks quantify systemic molecular aging but are blind to anatomy; imaging-based models capture structural aging but have no direct access to underlying regulatory state [29]. Recent organ-resolved approaches based on plasma proteomics [6] and UK Biobank phenotyping [5] have demonstrated that organ-specific aging rates predict disease and mortality, while multimodal image-based systems [15] confirm that fusing diverse imaging modalities improves biological age estimation. Organ-Age complements these efforts by combining molecular (transcriptomic) and structural (radiological) data in a single framework, yielding organ-age gaps that reflect structural *and* regulatory change. The per-organ metrics (Table 3) confirm that different organs exhibit distinct prediction profiles: brain achieves the tightest residuals (SD = 6.48 years) while lung shows the widest (SD = 17.29 years), a 2.7× difference that would be invisible in an organism-wide clock.

The organ-age deltas behave as structured residuals rather than pure error. Different organs show different dispersion profiles, and residual variance grows differently with age across tissues. This supports a practical framing: biological age is not a single scalar property uniformly expressed across the body. It is a distributed quantity with organ-specific trajectories shaped by exposure, repair capacity, and tissue-specific programs.

From a translational perspective, organ-level deltas offer a localized deviation signal that can complement organism-wide clocks. Exposures such as smoking and heavy alcohol use disproportionately affect lungs and liver. In principle, organ-resolved deltas could flag localized acceleration even when systemic biological age remains near expected. This is not a diagnosis and cannot be treated as one. It is a prioritization signal: an organ-specific deviation that warrants follow-up in settings where metadata and longitudinal outcomes are available.

Several limitations deserve mention. The three datasets do not overlap at the individual level, so alignment operates on representations rather than matched samples; this increases flexibility but complicates causal interpretation across modalities. Because no subject-level multimodal pairing exists, the contrastive alignment objective assumes biological correspondence based on shared age structure rather than direct sample matching. An ablation on truly matched multimodal subsets would provide stronger evidence for alignment benefits, but such paired transcriptomic–radiological datasets are not currently available at the scale required; future cohorts that collect molecular and imaging data from the same individuals would enable this validation. The framework currently covers adult aging only and does not model development. While the probabilistic head yields uncertainty estimates, separating epistemic from aleatoric uncertainty remains difficult in practice. The sample size imbalance across modalities is substantial (187,825 X-ray samples versus 563 MRI), which means that X-ray dominates the training signal in the fused model; the strong MRI-specific MAE (6.21 years) should be interpreted in light of the smaller and more homogeneous IXI cohort rather than taken as representative of MRI performance at scale.

Natural next steps include adding proteomics, metabolomics, and longitudinal clinical data as additional modalities; evaluating the model longitudinally to test whether organ-age deviations predict outcomes; and expanding interpretability analyses to connect latent factors to pathways and cell-type composition changes.

4 Conclusion

This study introduced Organ-Age, a multimodal representation learning framework integrating transcriptomic and radiological signals for organ-specific biological age estimation. Contrastive alignment combined with transformer-based fusion improved predictive performance relative to unimodal baselines, demonstrating potential benefits of multimodal integration. Per-subject panels and attribution summaries extend the framework beyond prediction, distilling organ-level deviations into readable reports and tracing them to candidate transcriptomic drivers. However, interpretation is constrained by heterogeneous data sources, lack of subject-level multimodal pairing, and limited clinical validation. Future work should emphasize longitudinal datasets, balanced modality sampling, and causal inference analyses to establish robustness and translational applicability.

References

[1] Steve Horvath. DNA methylation age of human tissues and cell types. *Genome Biology*, 14(10):3156, December 2013.

[2] Gregory Hannum, Justin Guinney, Lei Zhao, Lei Zhang, Gregory Hughes, Srinivas Sadda, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, 49(2):359–367, 2013.

[3] Morgan E. Levine, Ake T. Lu, Austin Quach, Brian H. Chen, Themistocles L. Assimes, Stefania Bandinelli, Lifang Hou, Andrea A. Baccarelli, James D. Stewart, Yun Li, Eric A. Whitsel, James G Wilson, Alex P Reiner, Abraham Aviv, Kurt Lohman, Yongmei Liu, Luigi Ferrucci, and Steve Horvath. An epigenetic biomarker of aging for

lifespan and healthspan. *Aging*, 10(4):573–591, April 2018.

[4] Enikő Zakar-Polyák, Attila Csordas, Róbert Pálovics, and Csaba Kerepesi. Profiling the transcriptomic age of single-cells in humans. *Communications Biology*, 7(1):1397, October 2024.

[5] Ye Ella Tian, Vanessa Cropley, Andrea B. Maier, Nicola T. Lautenschlager, Michael Breakspear, and Andrew Zalesky. Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality. *Nature Medicine*, 29:1221–1231, 2023.

[6] Hamilton Se-Hwee Oh, Jarod Rutledge, Daniel Nachun, Róbert Pálovics, Jordan Aber, GiGi McMahon, et al. Organ aging signatures in the plasma proteome track health and disease. *Nature*, 624:164–172, 2023.

[7] James H. Cole, Rudra P.K. Poudel, Dimosthenis Tsagkrasoulis, Matthan W.A. Caan, Claire Steves, Tim D. Spector, and Giovanni Montana. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, 163:115–124, December 2017.

[8] Pranav Rajpurkar, Jeremy Irvin, Robyn L. Ball, Kaylie Zhu, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Medicine*, 15(11):e1002686, November 2018.

[9] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation Learning with Contrastive Predictive Coding, 2018.

[10] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A Simple Framework for Contrastive Learning of Visual Representations, July 2020.

[11] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision, February 2021.

[12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need, August 2017.

[13] Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, et al. Perceiver IO: A General Architecture for Structured Inputs & Outputs, 2021.

[14] Amanpreet Singh, Ronghang Hu, Vedanuj Goswami, et al. FLAVA: A Foundational Language And Vision Alignment Model, 2021.

[15] Junhao Wang, Yong Gao, Fangji Wang, Simin Zeng, Jing Li, Huabin Miao, Ting Wang, Jiangchuan Zeng, Daniel T. Baptista-Hon, Oscar Monteiro, et al. Accurate estimation of biological age and its application in disease prediction using a multimodal image Transformer system. *Proceedings of the National Academy of Sciences*, 121(3):e2308812120, 2024.

[16] Alex Kendall and Yarin Gal. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?, 2017.

[17] GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature*, 550(7675):204–213, 2017.

[18] Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. *AAAI*, 33(01):590–597, 2019.

[19] IXI Dataset Consortium. Information extraction from images (ixi) dataset. https://brain-development.org/ixi-dataset/, 2020. Accessed 2025.

[20] Michael I. Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biology*, 15(12):550, 2014.

[21] W. Evan Johnson, Cheng Li, and Ariel Rabinovic. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics*, 8(1):118–127, 2007.

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition, December 2015.

[23] Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, et al. Big Transfer (BiT): General Visual Representation Learning. In *ECCV 2020*, volume 12350, pages 491–507. Springer, Cham, 2020.

[24] B Avants, C Epstein, M Grossman, and J Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, February 2008.

[25] Stephen M. Smith, Mark Jenkinson, Mark W. Woolrich, et al. Advances in functional and structural

MR image analysis and implementation as FSL. *NeuroImage*, 23:S208–S219, January 2004.

[26] Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. Deep Variational Information Bottleneck, 2016.

[27] Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, 2018.

[28] Tadas Baltrusaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE TPAMI*, 41(2):423–443, February 2019.

[29] Evgeny Putin, Polina Mamoshina, Alexander Aliper, Mikhail Korzinkin, Alexey Moskalev, Alexey Kolosov, Alexander Ostrovskiy, Charles Cantor, Jan Vijg, and Alex Zhavoronkov. Deep biomarkers of human aging: Application of deep neural networks to biomarker development. *Aging*, 8(5):1021–1033, May 2016.