This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Multimodal contrastive regression for organ-resolved biological age prediction

Veronika Ecker[1,2], Tim Förster[1], Sergios Gatidis[3], Thomas Küstner[2], and Bin Yang[1]

[1] Institute of Signal Processing and System Theory, University of Stuttgart
[2] Diagnostic and Interventional Radiology, University Hospital of Tuebingen
[3] Center for Artificial Intelligence in Medicine and Imaging, Stanford University
veronika.ecker@iss.uni-stuttgart.de

Abstract. Aging is complex and heterogeneous. A person's chronological age (CA) cannot reflect individual aging trajectories which prompted investigations for biological age (BA). Accurate estimation of BA therefore requires to account for the combined influence of structural, functional, and physiological factors on the aging process. BA progresses heterogeneously across organs, motivating models that integrate multiple data types to capture this complexity. Existing BA models however often rely on single organs or limited biomarkers, restricting holistic assessment. Overcoming these constraints requires multimodal models that better reflect the distributed nature of aging. We propose a multimodal contrastive regression framework that jointly learns from MRI and structured clinical variables to estimate organ-specific BA. A contrastive regression loss structures the latent space to reflect continuous age differences both within and across modalities. Applied to a large population cohort, this approach produces well-structured latent embeddings, improves age estimation relative to unimodal systems, and reveals characteristic age-gap patterns in healthy and diseased subgroups. The results demonstrate that combining MRI with complementary tabular features strengthens BA estimation and supports a comprehensive multi-organ view of aging.

Keywords: Biological Age · Multimodal Contrastive Regression

1 Introduction

The risk of developing many prevalent diseases, including cardiovascular disease [23, 26], cancer [8], and neurodegenerative disorders [20, 7], increases significantly with advancing age. Accurate assessment of an individual's aging process offers valuable insights into health status and disease risk. Conventionally, chronological age (CA) is used as the proxy for aging, but it only reflects time since birth and does not capture the complex interplay of genetic, environmental, and lifestyle factors that drive functional decline. In this regard, biological age (BA) reflects the functional state of organs and tissues beyond CA. However, estimating BA remains challenging due to the absence of a definitive ground truth and

the heterogeneous aging trajectories observed across different organ systems [33]. Prior efforts to estimate BA have relied primarily on molecular profiles [17, 14], laboratory measurements [24, 30], or functional tests [29, 25]. However, they capture only part of the aging process and overlook organ-specific variations.

The recent rise of deep learning has enabled to derive aging information directly from medical images, including Magnetic Resonance Imaging (MRI) [16, 10, 5], Computed Tomography (CT) [1], and X-ray [18]. Despite growing interest in organ-specific aging, no established organ-specific age predictors exist. Assessing BA therefore requires external validation using biological or clinical characteristics measured within the same individuals.

In the absence of such reference standards, and given the multifactorial nature of aging, we hypothesize that integrating complementary data sources may improve robustness and biological relevance of BA estimates. Previous multimodal learning studies paired structured clinical information with imaging inputs including MRI [13], CT [12], or chest radiographs [35]. Other works link tabular variables with time series signals like electrocardiographic data [32]. These approaches vary mainly in encoder design and fusion strategy and primarily focus on classification, while regression tasks remain underexplored.

In multimodal learning, self-supervised contrastive learning has emerged as a powerful pre-training strategy [13], grouping subjects with similar properties in the latent space. While this naturally supports classification, regression tasks require additional structure to reflect the ordering of continuous targets [9, 3].

Building on these advances, we propose a multimodal contrastive regression framework for organ-specific BA estimation. Our contribution is twofold: first, we extend imaging-based single-organ age estimation to a multiorgan framework. Second, we propose a multimodal contrastive regression model combining MRI with complementary tabular features, including lifestyle factors, blood-based measures, and imaging-derived biomarkers, to predict BA in six organ systems: brain, heart, liver, kidneys, spleen, and pancreas. Training on a healthy UK Biobank cohort using CA captures normative aging patterns, while testing is performed on a mixed subgroup. By jointly exploiting imaging and tabular data, our results highlight the value of combining complementary information for accurate, organ-specific BA assessment (code available at https://github.com/lab-midas/multimodal-age).

2 Materials and Methods

2.1 Data

The study uses the UK Biobank cohort, containing health information from adults aged 44 to 83 years across the UK, including extensive phenotypic information and imaging data. Available MRI scans include 3D MPRAGE brain, 2D+time cardiac cine, and 3D abdominal Dixon VIBE from ∼70,000 participants. Brain images were skull stripped and aligned to MNI152 space. [11]. After segmentation using two pre-trained segmentation models [2, 19], cardiac and abdominal images were cropped to extract organ-volumes of heart, liver, spleen,

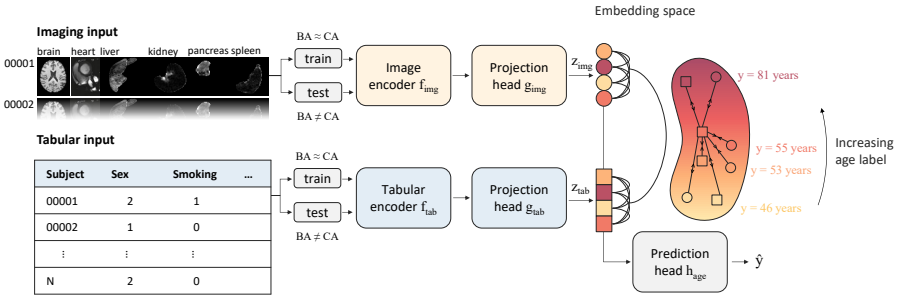

Fig. 1. Overview of the contrastive framework for organ-specific biological age (BA) estimation. Organ-specific imaging and tabular inputs are split into a healthy training cohort and a mixed (healthy and diseased) test cohort. Contrastive pre-training orders embeddings by age distance. After pre-training, a prediction head is trained using the ordered feature embeddings as input to predict the age $\hat{y}$. During training we use healthy controls where we can assume that BA $\approx$ chronological age (CA).

kidneys, and pancreas. For abdominal organs, the background was masked to prevent shortcut learning through fat tissue outside target structures. For organ-specific tabular data, we selected features linked to organ aging such as imaging-derived biomarkers as well as blood-based measurements and lifestyle factors. Specifically, variables were selected from organ-relevant domains, for example cognitive tests for the brain; blood pressure parameters for the heart; different blood biochemistry markers for the liver, spleen, and pancreas; urine biomarkers for the kidney. A full list of selected data fields is available in GitHub (https://github.com/lab-midas/multimodal-age). Continuous variables were z-score normalized, and cardinal categorical variables were one-hot encoded. For downstream analysis, disease onset ages were derived from ICD-10 codes with diagnosis dates and categorized as pre-existing or future relative to the imaging time point.

2.2 Multimodal integration

Our framework adopts the SimCLR [13, 4] as a backbone with separate encoders and projection heads for image and tabular data (Fig. 1). For each subject, imaging data x_{img} and tabular data x_{tab} are processed by their respective encoder and projection branches. The image encoder $f_{\text{img}}(\cdot)$ is based on a ResNet18 [15] modified to 3D data for leveraging spatial (brain, abdomen) or spatiotemporal context (heart). The tabular encoder $f_{\text{tab}}(\cdot)$ is a five-layer perceptron with batch normalization and ReLU activations. The resulting embeddings are mapped into a shared projection space through projection heads g_{img} and g_{tab}. Each projection head consists of two dense layers with layer normalization and LeakyReLU activation, yielding final projections $z_{\text{img}} = g_{\text{img}}(f_{\text{img}}(x_{\text{img}})) \in \mathbb{R}^{128}$ and $z_{\text{tab}} = g_{\text{tab}}(f_{\text{tab}}(x_{\text{tab}})) \in \mathbb{R}^{128}$. These projections are then optimized in

a shared embedding space to align the modalities according to their age label difference.

2.3 Multimodal contrastive regression loss

In classification-based contrastive learning, embeddings of positive pairs are pulled together while negative pairs are pushed apart. However, in regression tasks, the labels are continuous, requiring the embedding space to follow a continuous structure rather than a categorical scheme. Accordingly, pairwise attraction should scale with label proximity. To adapt the standard InfoNCE loss [27], we use a weighting factor based on a Gaussian kernel function $w_{i,j} = K(y_i - y_j)$ [9, 3], where y_i and y_j are the continuous labels of samples i and j, respectively. In a batch of size B, every sample j relative to the anchor i is initially treated as a potential positive, but the weighting factor reduces the influence of pairs with large label distance. Only samples with sufficiently similar age labels to y_i contribute meaningfully to the loss. This weighting scheme yields a contrastive regression objective enforcing continuous structure in latent space. The tolerance margin for label similarity is controlled by the standard deviation σ of the Gaussian kernel, which should be selected according to sample size, batch size, and label distribution. To extend this formulation to the multimodal setting, we apply the loss within and across modalities to align paired imaging and tabular embeddings while preserving label ordering. For an anchor i from modality $m \in \{1, 2, ..., M\}$, candidates from all modalities contribute to the loss term l_{i_m}:

$$l_{i_m} = \sum_{n=1}^{M} \sum_{j=1}^{B} \mathbb{1}_{j \neq i, n \neq m} \frac{w_{i,j}}{\sum_{j=1}^{B} \mathbb{1}_{j \neq i} w_{i,j}} \left(- \ln \frac{s_{i_m, j_n}}{\sum_{n=1}^{M} \sum_{j=1}^{B} \mathbb{1}_{j \neq i, n \neq m} s_{i_m, j_n}} \right),$$

where the similarity is defined as $s_{i_m, j_n} = \exp(sim(z_{i_m}, z_{j_n}))$ using the cosine similarity $sim(\cdot, \cdot)$. In our experimental setup, we consider two modalities ($M = 2$), yielding the final multimodal contrastive regression loss:

$$\mathcal{L} = \frac{1}{BM} \sum_{i=1}^{B} \sum_{m=1}^{M} l_{i_m}$$

2.4 Biological age prediction

Following contrastive pre-training, the learned ordered representations z_m, $m \in \{\text{img}, \text{tab}\}$, are concatenated and serve as input to the age prediction head (see Fig. 1). Unlike standard SimCLR, projection heads are retained, as the projections z_m preserve better age ordering than the raw embeddings. Therefore, the encoder and projection heads could, in principle, be merged. The age prediction head h_{age} is a four-layer perceptron with batch normalization and LeakyReLU activations, and a final linear layer to output the predicted age $\hat{y}_i$. During this stage, encoder and projection weights remain frozen and the age prediction head alone is optimized for age prediction using mean squared error

MSE $= \frac{1}{B}\sum_{i=1}^{N}|\hat{y}_i - y_i|$ loss with respect to the CA label. Since BA lacks direct ground truth, we approximate it using CA within a carefully selected organ-specific healthy cohort, assuming BA $\approx$ CA for this population. When applied to mixed healthy and diseased cohorts, the predicted age gap $\text{PAG}_i = \hat{y}_i - y_i$ is interpreted as accelerated or decelerated biological aging.

3 Experimental Setting

Separate models were trained for each organ. A healthy subcohort was defined by excluding participants with predefined organ-specific diagnoses from UK Biobank records, allowing CA to serve as a proxy label within the healthy cohort. The healthy cohort was split 70/30 into train and test sets. Due to organ-specific imaging availability and exclusion criteria, training sample sizes ranged from 19,128 (brain) to 30,641 (spleen). Data augmentation included resizing, cropping, mirroring, and gamma adjustment. Contrastive pre-training was conducted for 12 epochs on two A100 GPUs, followed by 10 epochs of training of the age prediction head (batch size of 100, Adam optimizer [22]). The Gaussian kernel parameter for the contrastive regression loss was selected using grid search. After training, bias correction was performed using linear regression to account for regression-to-the-mean effects.

In the absence of validated organ-specific BA predictors, no validated benchmarks exist. Since previous approaches often rely on traditional machine learning or shallow networks, comparisons with deeper models are potentially unfair. To provide an initial reference point for future external validation, we trained a unimodal imaging-based baseline using a ResNet architecture [16]. However, comparisons with this baseline should be interpreted with caution. To better assess the role of multimodal fusion, unimodal ablation models ($M = 1$) were also trained under the same settings as multimodal training. To gain further insights into tabular performance, the tabular features in the unimodal experiments were divided into imaging-derived biomarkers and non-imaging variables and separately assessed using mean absolute errors $\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|$.

In the absence of a direct ground truth and established organ-specific reference methods, the model validation necessarily relies on external clinical and biological metadata to assess whether prediction errors capture meaningful information. To evaluate the clinical significance of predicted age gaps, we considered both diagnostic and prognostic scenarios. We first compared average PAG between individuals with pre-existing conditions and healthy controls. In the prognostic analysis, left-truncated, right-censored Kaplan–Meier models assessed associations between PAG and time-to-disease.

4 Results and Discussion

Pre-training was mostly effective across all organs with few exceptions of spleen and pancreas (Fig. 2). The loss enforced CA-ordered alignment within and across imaging and tabular modalities. Tabular encoders were robust across organs,

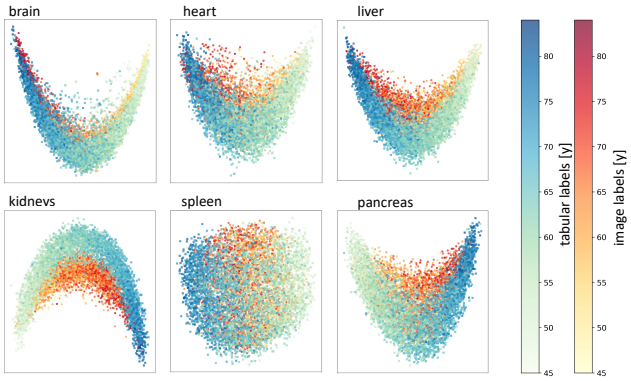

Fig. 2. Visualization of global patterns in imaging and tabular feature embeddings after contrastive pre-training for all organs using principal component analysis plots. Color grading shows age labels for imaging (yellow–red) and tabular (green–blue) features.

whereas image encoders depended on the strength of structural aging signals. In the brain, strong markers of aging, including volumetric reductions in gray and white matter [21], produced well-structured embeddings with few outliers. In comparison, pancreas and especially spleen showed more outliers, consistent with weaker visible aging markers. While Fig. 2 indicates coherent intra-modal ordering and broad multimodal alignment, it does not demonstrate explicit individual-level alignment. After training, performance was assessed on a healthy test set as lower MAE in mixed cohorts do not necessarily reflect better prediction. MAE for the baseline ResNet models, multimodal models and unimodal ablations are summarized in Fig. 3. Multimodal training effectively leveraged complementary information in all organs, surpassing all unimodal models and baseline in all organs. The ResNet baseline showed performance comparable to the unimodal image model. Because no BA ground truth exists, comparisons should be interpreted cautiously. Raw image and image-biomarker unimodal models worked well in organs where aging produces clear structural changes, such as the brain, but were less effective in organs with limited age-related imaging signals (kidney, spleen, and pancreas). In organs with strong imaging markers of aging, raw images supported better predictions. In organs with weak aging signals (spleen and kidney) handcrafted image biomarkers performed better than raw images. Tabular-only models were consistent across organs and generally outperformed imaging-only models, except in the brain where imaging information was particularly informative.

By including participants with diseases, PAG can be interpreted as an indicator of BA, providing insight into accelerated or decelerated organ aging. Diseased subgroups were selected based on established links between organ aging and common conditions (brain - Alzheimer's disease [6, 5], pancreas - type II diabetes mellitus [31], kidney - chronic kidney disease [28, 34]). Consistent with literature, average PAG appeared elevated in certain diseased subgroups comp-

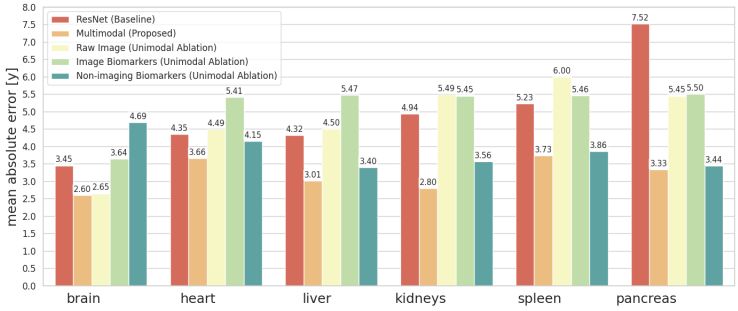

Fig. 3. Mean absolute errors of predicted ages in healthy test set for multimodal trainings and unimodal ablations.

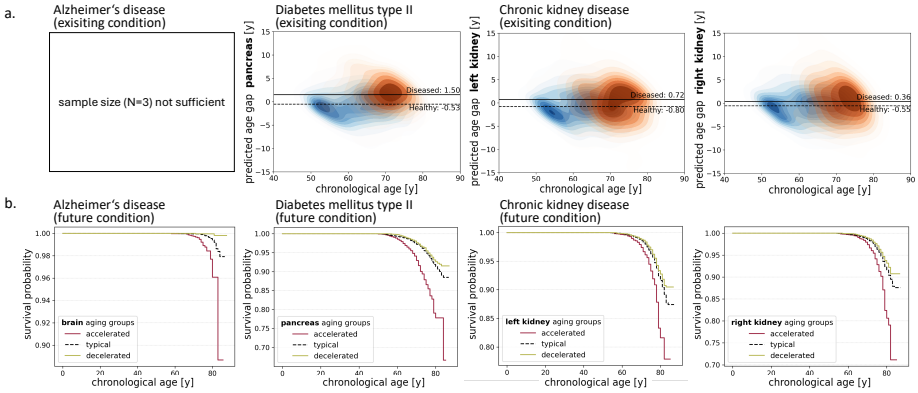

Fig. 4. a. Kernel density estimations of PAG (= predicted age - chronological age) across organs for specific health outcomes, comparing subjects with pre-existing diagnoses (orange) with healthy controls (blue). Mean values of PAG are shown for the healthy cohort (dashed line) and diseased cohort (solid line). b. Kaplan-Meier survival curves show time-to-event for accelerated (PAG $>$ +SD), normative, and decelerated (PAG $<$ −SD) aging groups (SD = organ-specific standard deviation of PAG).

ared with healthy controls (Fig. 4a). Welch's t-tests confirmed significant differences ($p < 1.67 \times 10^{-2}$; $N = 3$, Bonferroni-corrected) in mean PAG between healthy and diseased groups (type II diabetes - pancreas: $d = 0.59$; chronic kidney disease - left kidney: $d = 0.27$; right kidney: $d = 0.30$), suggesting that PAG may reflect disease-related deviations in organ aging. To assess the prognostic value of PAG, Kaplan–Meier analyses compared time-to-disease across accelerated (PAG $>$ +SD), normative, and decelerated (PAG $<$ −SD) aging groups (SD: organ-specific PAG standard deviation). Accelerated aging consistently showed shorter time to onset (Fig. 4b), while differences between normative and decelerated groups were smaller but most pronounced for Alzheimer's disease,

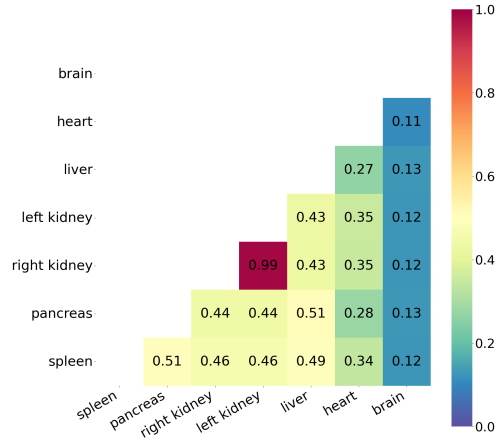

Fig. 5. Heatmap of Pearson correlation coefficients between predicted age gaps of brain, heart, left kidney, right kidney, pancreas, and spleen. All results are statistically significant with $p < 2.38 \times 10^{-3}$ (21 tests, Bonferroni-corrected).

where decelerated brain aging had the highest event-free survival. Univariate Cox regression confirmed significant differences ($p < 1.25 \times 10^{-2}$, $N = 4$, Bonferroni-corrected) in hazard ratios (HR) for Alzheimer's disease (HR = 4.56), type II diabetes mellitus (HR = 1.58), and chronic kidney disease (left: HR = 1.27, right: HR = 1.30).

Beyond disease associations, cross-organ correlations were examined to understand systemic relationships (Fig. 5). Brain aging occurs relatively independently, whereas abdominal organs and the heart are moderately correlated. Strong correlations between the left and right kidneys reflect expected behavior of bilateral organs. Overall, organs age heterogeneously but shared patterns are visible across the system.

We acknowledge several limitations. Our framework assumes BA $\approx$ CA in a diagnosis excluded healthy cohort, although undiagnosed or unreported conditions may still influence training. Moreover, normative aging is not defined solely by the absence of disease, especially in older populations. Without ground truth BA labels, this assumption remains a practical approximation currently used in the field, supported by comprehensive UK Biobank National Health Service records. Additionally, the older demographic of the cohort may limit generalization to younger populations and immediate clinical applicability, motivating future validation in the younger German National Cohort. External validation is further restricted by the lack of organ-specific BA benchmarks. Epigenetic clocks [17, 14] are well recognized but lack organ specificity. We therefore rely on examining aging related associations and plan phenome-wide association studies to strengthen evaluation. Finally, limited age-related imaging signals in some organs may impair training performance, which could be mitigated through advanced fusion techniques in future work.

5 Conclusion

In summary, our study demonstrates that integrating complementary aging information from MRI and structured health data through contrastive regression benefits organ-specific BA estimation across six organs. Multimodal models outperform unimodal approaches, detect accelerated aging in diseased individuals, and reveal cross organ aging patterns.

Acknowledgments. This work was carried out under UK Biobank Application 60520. We thank all participants who took part in the UKBB study and the staff in this research program. The work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project #428219130 and supported under Germany's Excellence Strategy (EXC 2064/1, #390727645).

Disclosure of Interests. The authors declare no competing financial or non-financial interests.

References

1. Azarfar, G., et al.: Deep learning-based age estimation from chest CT scans **19**(1), 119–127 (2024). https://doi.org/10.1007/s11548-023-02989-w
2. Bai, W., et al.: Automated cardiovascular magnetic resonance image analysis with fully convolutional networks **20**(1), 65 (2018). https://doi.org/10.1186/s12968-018-0471-x
3. Barbano, C.A., et al.: Contrastive learning for regression in multi-site brain age prediction. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). pp. 1–4 (2023). https://doi.org/10.1109/ISBI53787.2023.10230733
4. Chen, T., et al.: A simple framework for contrastive learning of visual representations. In: Proceedings of the 37th International Conference on Machine Learning. ICML'20, JMLR.org (2020)
5. Cole, J.H., et al.: Brain age predicts mortality **23**(5), 1385–1392 (2018). https://doi.org/10.1038/mp.2017.62
6. Cole, J.H., Franke, K.: Predicting age using neuroimaging: Innovative brain ageing biomarkers **40**(12), 681–690 (2017). https://doi.org/10.1016/j.tins.2017.10.001
7. Corrada, M., et al.: Sources of variability in prevalence rates of alzheimer's disease **24**(5), 1000–1005 (1995). https://doi.org/10.1093/ije/24.5.1000
8. DePinho, R.A.: The age of cancer **408**(6809), 248–254 (2000). https://doi.org/10.1038/35041694
9. Dufumier, B., et al.: Contrastive learning with continuous proxy meta-data for 3d MRI classification. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. pp. 58–68. Springer International Publishing (2021)
10. Ecker, V.e.a.: Deep regression for biological age estimation in multiple organs: Investigations on 40,000 subjects of the UK biobank. In: ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 2255–2259 (2024). https://doi.org/10.1109/ICASSP48485.2024.10447990
11. Evans, A., et al.: 3d statistical neuroanatomical models from 305 MRI volumes. In: 1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference. pp. 1813–1817 vol.3 (1993). https://doi.org/10.1109/NSSMIC.1993.373602

12. Fang, C., et al.: Deep learning for predicting COVID-19 malignant progression **72**, 102096 (2021). https://doi.org/10.1016/j.media.2021.102096
13. Hager, P.e.a.: Best of both worlds: Multimodal contrastive learning with tabular and imaging data. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 23924–23935 (2023). https://doi.org/10.1109/CVPR52729.2023.02291
14. Hannum, G., et al.: Genome-wide methylation profiles reveal quantitative views of human aging rates **49**(2), 359–367 (2013). https://doi.org/10.1016/j.molcel.2012.10.016
15. He, K., et al.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90, ISSN: 1063-6919
16. Hepp, T., et al.: Uncertainty estimation and explainability in deep learning-based age estimation of the human brain: Results from the german national cohort MRI study **92**, 101967 (2021). https://doi.org/10.1016/j.compmedimag.2021.101967
17. Horvath, S.: DNA methylation age of human tissues and cell types **14**(10), 3156 (2013). https://doi.org/10.1186/gb-2013-14-10-r115
18. Ieki, H., et al.: Deep learning-based age estimation from chest x-rays indicates cardiovascular prognosis **2**(1), 1–12 (2022). https://doi.org/10.1038/s43856-022-00220-6, publisher: Nature Publishing Group
19. Kart, T., et al.: Deep learning-based automated abdominal organ segmentation in the UK biobank and german national cohort magnetic resonance imaging studies **56**(6), 401–408 (2021). https://doi.org/10.1097/RLI.0000000000000755
20. Katzman, R., Saitoh, T.: Advances in alzheimer's disease **5**(3), 278–286 (1991). https://doi.org/10.1096/fasebj.5.3.2001787
21. Kaye, J.A., et al.: The significance of age-related enlargement of the cerebral ventricles in healthy men and women measured by quantitative computed x-ray tomography **40**(3), 225–231 (1992). https://doi.org/10.1111/j.1532-5415.1992.tb02073.x
22. Kingma, D., Ba, J.: Adam: A method for stochastic optimization (2014)
23. Lakatta, E.G.: Age-associated cardiovascular changes in health: Impact on cardiovascular disease in older persons **7**(1), 29–49 (2002). https://doi.org/10.1023/A:1013797722156
24. Le Goallec, A., Patel, C.J.: Age-dependent co-dependency structure of biomarkers in the general population of the united states **11**(5), 1404–1426 (2019). https://doi.org/10.18632/aging.101842
25. Nakamura, E., et al.: Biological age versus physical fitness age in women **61**(3), 202–208 (1990). https://doi.org/10.1007/BF00357600
26. North, B.J., Sinclair, D.A.: The intersection between aging and cardiovascular disease **110**(8), 1097–1108 (2012). https://doi.org/10.1161/CIRCRESAHA.111.246876
27. Oord, A.v.d., et al.: Representation learning with contrastive predictive coding. **abs/1807.03748** (2018)
28. O'Sullivan, E.D., et al.: Renal aging: Causes and consequences **28**(2), 407–420 (2017)
29. Rahman, S.A., Adjeroh, D.A.: Deep learning using convolutional LSTM estimates biological age from physical activity **9**(1), 11425 (2019). https://doi.org/10.1038/s41598-019-46850-0, publisher: Nature Publishing Group
30. Sagers, L., et al.: Prediction of chronological and biological age from laboratory data **12**(9), 7626–7638 (2020). https://doi.org/10.18632/aging.102900

31. Shen, B., et al.: Association between age at diabetes onset or diabetes duration and subsequent risk of pancreatic cancer: results from a longitudinal cohort and mendelian randomization study **30** (2023). https://doi.org/10.1016/j.lanwpc.2022.100596
32. Stym-Popper, J., et al.: DAFTED: Decoupled asymmetric fusion of tabular and echocardiographic data for cardiac hypertension diagnosis. In: Proceedings of Machine Learning Research (2025)
33. Tian, Y.E., et al.: Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality **29**(5), 1221–1231 (2023). https://doi.org/10.1038/s41591-023-02296-6
34. Vlassara, H., et al.: Role of oxidants/inflammation in declining renal function in chronic kidney disease and normal aging (114), S3–11 (2009). https://doi.org/10.1038/ki.2009.401
35. Zhou, H.Y., et al.: A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics **7**(6), 743–755 (2023). https://doi.org/10.1038/s41551-023-01045-x