FaceTTD: Time-to-Death from Imageomic Facial Time Series for Rapid Mortality Risk Profiling and Longevity Interventions

Saleem A. Al Dajani [*1,2,3] **Dmitrii Glubokov**[2] **Alexander Tyshkovskiy**[2] **John R. Williams**[1]
Jesse Poganik[2] **Omar Abudayyeh**[*,2] **Jonathan Gootenberg**[*,3] **Vadim N. Gladyshev**[*,2]

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology
[2]Brigham and Women's Hospital, Mass General Brigham, Harvard Medical School
[3]Beth Israel Deaconess Medical Center, Beth Israel Lahey Health, Harvard Medical School

Abstract

Mortality risk assessment remains a fundamental challenge in healthcare, with current methods relying on chronological age that fails to capture individual variation in biological aging and proximity to death. Direct prediction of time-to-death from accessible, non-invasive phenotypic signals could enable more precise aging risk stratification and targeted longevity interventions. We present **FaceTTD**, a framework for predicting time-to-death (TTD) from facial images as a measure of biological aging. Treating portraits as time series inputs, we train XGBoost and Random Forest regressors on curated IMDB and Wikipedia datasets. In-distribution performance reaches $R^2 = 0.67$, but falls to $R^2 = 0.12$–0.25 out-of-distribution depending on TTD subset. Longitudinal facial trajectories improve predictive accuracy, indicating value in temporal coverage. Our findings highlight the promise and limitations of mortality modeling from phenotypic time series, positioning mortality horizon estimation as an imageomics problem where facial trajectories serve as accessible phenotypes of biological aging, and motivating multimodal extensions (voice, video, wearables, EHRs) for robust health applications. A live demo is available at `https://huggingface.co/spaces/doubleblindanonymous/facettd`.

1 Introduction

Accurate mortality risk assessment is fundamental to healthcare resource allocation, clinical decision-making, and preventive interventions, yet current approaches fail to capture the heterogeneity of biological aging and individual proximity to death. While extensive research has focused on predicting chronological age from biomarkers and phenotypic features [1–11], these methods inherently assume uniform aging trajectories and cannot distinguish between individuals of the same age who may have vastly different mortality horizons. The inability to directly model *time-to-death* (TTD) from accessible, non-invasive signals represents a critical gap in precision medicine, limiting our capacity to identify high-risk individuals who would benefit most from targeted interventions.

Facial characteristics encode cumulative biological patterns that manifest as visible aging markers, providing an accessible window into biological rather than chronological age. Unlike single-timepoint assessments, longitudinal facial changes capture the dynamic progression of biological aging, revealing accelerated or decelerated aging trajectories that correlate with mortality risk [12–14]. TTD provides a more direct outcome, aligning with chronic disease detection and intrinsic capacity as defined by the World Health Organization [4, 15].

We investigate TTD prediction as an initial mortality proxy from facial images, treating portraits across time as phenotypic time series. By learning from temporal sequences rather than isolated snapshots, we aim to capture mortality-relevant phenotypic changes that static age estimation cannot detect. We evaluate in-distribution (ID) vs. out-of-distribution (OOD) performance and explore the role of longitudinal coverage in building robust mortality risk models from facial time series [16, 17].

[*]Corresponding authors: `sdajani@mit.edu, oabudayyeh@bwh.harvard.edu, jgootenb@bidmc.harvard.edu, vgladyshev@rics.bwh.harvard.edu`

39th Conference on Neural Information Processing Systems (NeurIPS 2025), The 3rd Workshop on Imageomics: Discovering Biological Knowledge from Images Using AI.

2 Data and Methods

Two TTD-labeled datasets were constructed: (1) *IMDB* (in-distribution), pairing celebrity portraits from the IMDB-Face corpus [18, 19] with mortality metadata from the IMDB database; and (2) *Wiki* (out-of-distribution), built by extracting Wikipedia biographies with recorded birth/death dates and linking to age-annotated images in the WIKI face set [18, 19]. TTD was computed as the difference between photo timestamp and death date. Resized 64x64 grayscale facial image pixels were vectorized along with chronological age and input to XGBoost [20] and Random Forest [21] regressors. Hyperparameters were tuned via grid search. Performance was evaluated on in-distribution test sets and out-of-distribution (OOD) subsets stratified by TTD horizon. A complete description of the training and evaluation procedures can be found in Appendices A.1 and A.2.

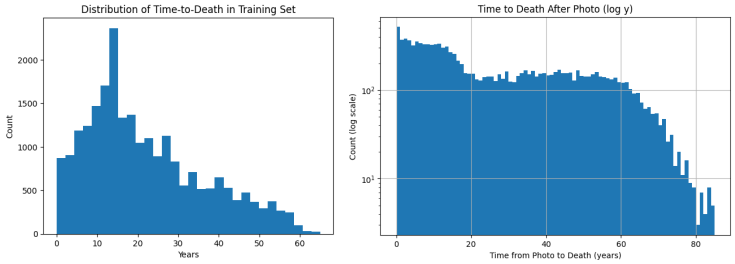

Figure 1: Distribution of time-to-death (TTD) in IMDB (left, denser near 0–40 years) and Wikipedia (right, broader coverage).

3 Results

On the IMDB-FaceTTD dataset, models achieve strong in-distribution performance. XGBoost reaches $R^2 = 0.67$ and MAE $= 4.1$ years on the held-out test set, while training accuracy is near-perfect ($R^2 = 1.0$, MAE < 0.1). Performance improves further when longitudinal coverage is available: subjects with multiple portraits over time yield higher R^2 values and lower MAEs than single-image baselines (Fig. 2, Fig. 3). These results establish that facial time series contain measurable signal for TTD inference, with temporal information contributing additional gains beyond static portraits.

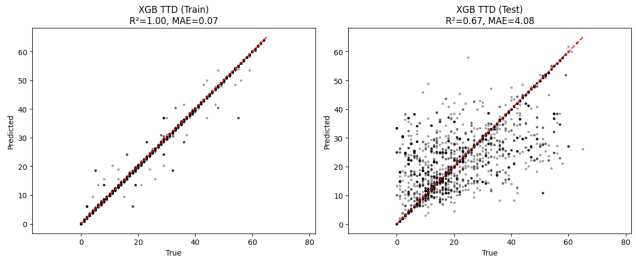

Figure 2: XGBoost performance on IMDB-FaceTTD train and test splits.

Out-of-distribution evaluation on Wiki-FaceTTD shows substantially weaker results. Here, R^2 drops to 0.12–0.25 depending on the TTD subset and regression method, and predictive slopes attenuate near death (Fig. 4). The in–out distribution gap exhibits a non-monotonic pattern: OOD performance improves moderately for mid-range horizons (TTD $<$ 50–60) but declines again for shorter horizons (TTD < 45), suggesting a near-term 'sweet spot' where TTD distributions align more closely (see Fig. 1). A complete breakdown of metrics is provided in Tab. 1.

4 Discussion

Our results reveal a critical challenge in mortality modeling: while FaceTTD achieves $R^2 = 0.67$ in-distribution, performance drops to $R^2 = 0.12 - 0.25$ out-of-distribution, exposing the brittleness

of facial aging markers between populations [22]. This generalization gap—from 82% to 40% predictive slope from ID to OOD—suggests our model captures more population-specific signatures than universal mortality biomarkers.

Our results show that while TTD prediction from facial images is feasible in-distribution, OOD generalization remains weak. The non-monotonic OOD performance, improving for TTD<50-60 years before declining at TTD<45, reveals a 'sweet spot' where facial mortality markers become more universal—potentially reflecting conserved end-of-life biological processes [9–11] that transcend demographics.

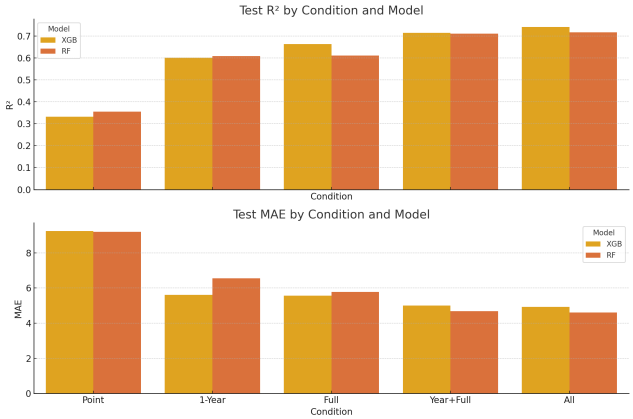

Figure 3: Adding longitudinal portraits improves performance.

Longitudinal coverage improves robustness, but accuracy still collapses under demographic and contextual shifts. Our finding that temporal sequences outperform single timepoints indicates that mortality risk is encoded in the rate of facial change, not static appearance. This suggests that clinical deployment would require repeated measurements to capture aging velocity [1, 3], not one-time assessments.

Table 1: Performance summary across TTD subsets by mean average error (MAE) in years. R^2 and $\sqrt{R^2}$ represent model fit and predictive slope, respectively.

Subset	Category	R^2	MAE (years)	$\sqrt{R^2}$
In-distribution (Train/Test)				
Train	XGBoost	1.00	0.07	1.00
Test	XGBoost	0.67	4.08	0.82
Out-of-distribution (OOD)				
OOD TTD	Optimized XGBoost	0.12	15.04	0.35
	Random Forest	0.19	14.63	0.44
	XGBoost	0.04	15.66	0.20
TTD $\leq$ 60	Optimized XGBoost	0.19	13.09	0.44
	Random Forest	0.25	12.80	0.50
	XGBoost	0.09	13.76	0.30
TTD $\leq$ 50	Optimized XGBoost	0.17	11.10	0.41
	Random Forest	0.20	10.99	0.45
	XGBoost	0.04	11.84	0.20
TTD $\leq$ 45	Optimized XGBoost	0.09	10.34	0.30
	Random Forest	0.10	10.36	0.32
	XGBoost	0.08	11.70	0.28
5 $\leq$ TTD $\leq$ 45	Optimized XGBoost	0.12	9.18	0.35
	Random Forest	0.18	8.94	0.42
	XGBoost	0.05	9.92	0.22

Improvements include filtering unnatural deaths, refining cause-of-death labels, and balancing demographic diversity. Multimodal time series [23] (voice [24–26], video [27–29], wearables [30–34], proteomic [35], heart [36], medical records [37–41]) may further enhance generalization. Combining facial trajectories with these complementary signals could better triangulate mortality risk across biological systems and may achieve the robustness needed for clinical deployment and impact.

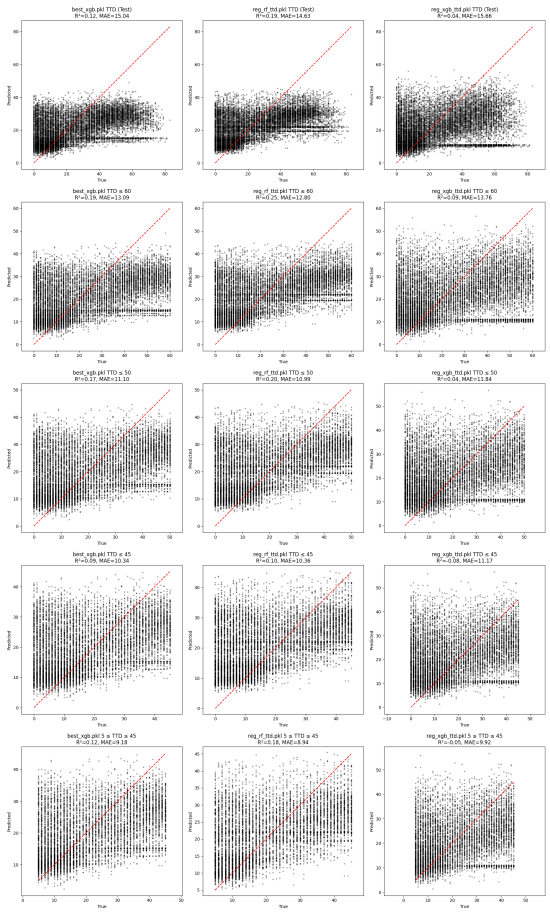

Figure 4: Predicted vs. true TTD scatter plots across OOD full Wiki dataset and TTD subsets. Performance degrades as death nears.

4.1 Limitations, Broader Impacts, and Safeguards

Our datasets are celebrity- and Wikipedia-biased, with uneven demographic representation. TTD labels depend on reported death dates and are used as a proxy for mortality. Experiments are limited to tree-based regressors; deep [2, 24] and foundation [6] models may capture richer signals but may also risk overfitting. Face-based mortality modeling raises concerns of bias, privacy, and misuse (e.g., surveillance). Responsible deployment requires safeguards, transparency, and strong ethical oversight. A usage disclaimer is included on the HuggingFace public demo as a safeguard, stating that it is intended solely for research in order to demonstrate the feasibility of phenotypic time-series modeling. This study highlights the intrinsic challenges of the facial imageomic phenotypic time-series problem while establishing a quantitative and systematic framework for analysis and evaluation.

5 Conclusion

We introduce FaceTTD, a framework for predicting TTD from facial time series. While models achieve $\sim$80% predictability in-distribution, OOD performance falls to $\sim$40%. These findings emphasize both the potential and limitations of face-based TTD modeling, and motivate multimodal, demographically balanced datasets for health applications, such as mortality risk profiling and nominating longevity interventions [42–44] based on a mechanistic understanding of aging [45, 46]. As a proof-of-concept, the study is constrained by confounders in the IMDB–Wikipedia imagery, and the single TTD target represents a preliminary mortality proxy. These results demonstrate the feasibility of future studies aimed at clinical applicability, framing longitudinal facial trajectories as a novel imageomic time-series paradigm for the digital biomarkers of aging.

NeurIPS Code of Ethics Statement

Research conducted in the paper conforms, in every respect, with the NeurIPS Code of Ethics [47].

Declaration of LLM Usage

This paper complies with the NeurIPS 2025 Policy on the Use of Large Language Models [48]. The extent of LLM involvement is detailed in Item 16 of the NeurIPS Paper Checklist (see Appendix).

Acknowledgments and Disclosure of Funding

This work was supported by grants from the NIA and the Hevolution Foundation. S.A.A. was funded by the MIT Department of Civil and Environmental Engineering (CEE) under the Friesecke (1961) Fellowship Fund. J.S.G. and O.O.A. are supported by NIH grants R01-EB031957, R01-AG074932, and R01-GM148745; G. Harold & Leila Y. Mathers Charitable Foundation; Rett Syndrome Research Trust; The Gordon and Betty Moore Foundation; Impetus Grants; Cystic Fibrosis Foundation Pioneer Grant; Google Deepmind; Sanofi; Yosemite; Michelson Foundation; Hevolution Foundation; American Federation for Aging Research; Pivotal Life Sciences; and the MGB Gene and Cell Therapy Institute

References

[1] Bence Király, Iván Fejes, and Csaba Kerepesi. Face photo-based age acceleration predicts all-cause mortality and differs among occupations. *bioRxiv*, 2025.

[2] Dennis Bontempi, Osbert Zalay, Danielle S Bitterman, Nicolai Birkbak, Derek Shyr, Fridolin Haugg, Jack M Qian, Hannah Roberts, Subha Perni, Vasco Prudente, et al. FaceAge, a deep learning system to estimate biological age from face photographs to improve prognostication: a model development and validation study. *The Lancet Digital Health*, 2025.

[3] Csaba Kerepesi and Botond Bárdos-Deák. Improving face age prediction by using multiple-angle photos. *bioRxiv*, pages 2025–08, 2025.

[4] Matías Fuentealba, Laure Rouch, Sophie Guyonnet, Jean-Marc Lemaitre, Philipe de Souto Barreto, Bruno Vellas, Sandrine Andrieu, and David Furman. A blood-based epigenetic clock for intrinsic capacity predicts mortality and is associated with clinical, immunological and lifestyle factors. *Nature Aging*, pages 1–10, 2025.

[5] M Austin Argentieri, Sihao Xiao, Derrick Bennett, Laura Winchester, Alejo J Nevado-Holgado, Upamanyu Ghose, Ashwag Albukhari, Pang Yao, Mohsen Mazidi, Jun Lv, et al. Proteomic aging clock predicts mortality and risk of common age-related diseases in diverse populations. *Nature Medicine*, 30(9):2450–2460, 2024.

[6] Fridolin Haugg, Grace Lee, John He, Leonard Nürnberg, Dennis Bontempi, Danielle S Bitterman, Paul Catalano, Vasco Prudente, Dmitrii Glubokov, Andrew Warrington, et al. Foundation Artificial Intelligence Models for Health Recognition Using Face Photographs (FAHR-Face). *arXiv preprint arXiv:2506.14909*, 2025.

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

[8] 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, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*, 10(4):573, 2018.

[9] Ake T Lu, Austin Quach, James G Wilson, Alex P Reiner, Abraham Aviv, Kenneth Raj, Lifang Hou, Andrea A Baccarelli, Yun Li, James D Stewart, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)*, 11(2):303, 2019.

[10] Ake T Lu, Alexandra M Binder, Joshua Zhang, Qi Yan, Alex P Reiner, Simon R Cox, Janie Corley, Sarah E Harris, Pei-Lun Kuo, Ann Z Moore, et al. DNA methylation GrimAge version 2. *Aging (Albany NY)*, 14(23):9484, 2022.

[11] Timothy V Pyrkov, Konstantin Avchaciov, Andrei E Tarkhov, Leonid I Menshikov, Andrei V Gudkov, and Peter O Fedichev. Longitudinal analysis of blood markers reveals progressive loss of resilience and predicts human lifespan limit. *Nature Communications*, 12(1):2765, 2021.

[12] Andreas Lanitis, Christopher J. Taylor, and Timothy F Cootes. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):442–455, 2002.

[13] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5810–5818, 2017.

[14] Zongwei Wang, Xu Tang, Weixin Luo, and Shenghua Gao. Face aging with identity-preserved conditional generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7939–7947, 2018.

[15] World Health Organization. *World report on ageing and health*. World Health Organization, 2015.

[16] Zachary C Lipton, David C Kale, Randall Wetzel, et al. Modeling missing data in clinical time series with rnns. *Machine Learning for Healthcare*, 56(56):253–270, 2016.

[17] Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M Dai, Nissan Hajaj, Michaela Hardt, Peter J Liu, Xiaobing Liu, Jake Marcus, Mimi Sun, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, 1(1):18, 2018.

[18] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Dex: Deep expectation of apparent age from a single image. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*, December 2015.

[19] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 126(2-4):144–157, 2018.

[20] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016.

[21] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[22] Brendan F Klare, Mark J Burge, Joshua C Klontz, Richard W Vorder Bruegge, and Anil K Jain. Face recognition performance: Role of demographic information. *IEEE Transactions on information forensics and security*, 7(6):1789–1801, 2012.

[23] Xiyuan Zhang, Boran Han, Haoyang Fang, Abdul Fatir Ansari, Shuai Zhang, Danielle C Maddix, Cuixiong Hu, Andrew Gordon Wilson, Michael W Mahoney, Hao Wang, et al. Does Multimodality Lead to Better Time Series Forecasting? *arXiv preprint arXiv:2506.21611*, 2025.

[24] David Krongauz, Hido Pinto, Sarah Kohn, Yanir Marmor, and Eran Segal. HPP-Voice: A Large-Scale Evaluation of Speech Embeddings for Multi-Phenotypic Classification. *arXiv: 2505.16490*, 2025.

[25] Felix Burkhardt, Johannes Wagner, Hagen Wierstorf, Florian Eyben, and Björn Schuller. Speech-based age and gender prediction with transformers. In *Speech Communication; 15th ITG Conference*, pages 46–50. VDE, 2023.

[26] Mingbin Xu, Alex Jin, Sicheng Wang, Mu Su, Tim Ng, Henry Mason, Shiyi Han, Zhihong Lei, Yaqiao Deng, Zhen Huang, et al. Conformer-based speech recognition on extreme edge-computing devices. *arXiv: 2312.10359*, 2023.

[27] Zitong Yu, Xiaobai Li, and Guoying Zhao. Facial-video-based physiological signal measurement: Recent advances and affective applications. *IEEE Signal Processing Magazine*, 38(6):50–58, 2021.

[28] Zitong Yu, Yuming Shen, Jingang Shi, Hengshuang Zhao, Philip HS Torr, and Guoying Zhao. Physformer: Facial video-based physiological measurement with temporal difference transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4186–4196, 2022.

[29] Jiankai Tang, Kequan Chen, Yuntao Wang, Yuanchun Shi, Shwetak Patel, Daniel McDuff, and Xin Liu. Mmpd: Multi-domain mobile video physiology dataset. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1–5. IEEE, 2023.

[30] Timothy V Pyrkov, Ilya S Sokolov, and Peter O Fedichev. Deep longitudinal phenotyping of wearable sensor data reveals independent markers of longevity, stress, and resilience. *Aging (Albany NY)*, 13(6):7900, 2021.

[31] Tim Althoff, Rok Sosič, Jennifer L Hicks, Abby C King, Scott L Delp, and Jure Leskovec. Large-scale physical activity data reveal worldwide activity inequality. *Nature*, 547(7663):336–339, 2017.

[32] Yuanyuan Chen, Shing Chan, Derrick Bennett, Xiaofang Chen, Xianping Wu, Yalei Ke, Jun Lv, Dianjianyi Sun, Lang Pan, Pei Pei, et al. Device-measured movement behaviours in over 20,000 china kadoorie biobank participants. *International Journal of Behavioral Nutrition and Physical Activity*, 20(1):138, 2023.

[33] Jinjoo Shim, Elgar Fleisch, and Filipe Barata. Wearable-based accelerometer activity profile as digital biomarker of inflammation, biological age, and mortality using hierarchical clustering analysis in nhanes 2011–2014. *Scientific Reports*, 13(1):9326, 2023.

[34] Eray Erturk, Fahad Kamran, Salar Abbaspourazad, Sean Jewell, Harsh Sharma, Yujie Li, Sinead Williamson, Nicholas J Foti, and Joseph Futoma. Beyond Sensor Data: Foundation Models of Behavioral Data from Wearables Improve Health Predictions. *arXiv: 2507.00191*, 2025.

[35] Hamilton Se-Hwee Oh, Yann Le Guen, Nimrod Rappoport, Deniz Yagmur Urey, Amelia Farinas, Jarod Rutledge, Divya Channappa, Anthony D Wagner, Elizabeth Mormino, Anne Brunet, et al. Plasma proteomics links brain and immune system aging with healthspan and longevity. *Nature Medicine*, pages 1–9, 2025.

[36] James Truslow, Angela Spillane, Huiming Lin, Katherine Cyr, Adeeti Ullal, Edith Arnold, Ron Huang, Laura Rhodes, Jennifer Block, Jamie Stark, et al. Understanding activity and physiology at scale: The apple heart & movement study. *npj Digital Medicine*, 7(1):242, 2024.

[37] Sarah Kohn, Alon Diament, Anastasia Godneva, Raja Dhir, Adina Weinberger, Yotam Reisner, Hagai Rossman, and Eran Segal. Phenome-wide associations of sleep characteristics in the Human Phenotype Project. *Nature Medicine*, 31(3):1026–1037, 2025.

[38] Lee Reicher, Smadar Shilo, Anastasia Godneva, Guy Lutsker, Liron Zahavi, Saar Shoer, David Krongauz, Michal Rein, Sarah Kohn, Tomer Segev, et al. Deep phenotyping of health–disease continuum in the Human Phenotype Project. *Nature Medicine*, pages 1–13, 2025.

[39] Yukang Jiang, Bingxin Zhao, Xiaopu Wang, Borui Tang, Huiyang Peng, Zidan Luo, Yue Shen, Zheng Wang, Zhiwen Jiang, Jie Wang, et al. UKB-MDRMF: a multi-disease risk and multimorbidity framework based on UK biobank data. *Nature Communications*, 16(1):3767, 2025.

[40] The UK Biobank Whole-Genome Sequencing Consortium. Whole-genome sequencing of 490,640 uk biobank participants. *Nature*, August 2025.

[41] Varun Dwaraka, Qingwen Chen, Natalia Carreras-Gallo, Kevin Mendez, Yulu Chen, Sofina Begum, Priyadarshini Kachroo, Nicole Prince, Hannah Went, Tavis Mendez, et al. Omicmage: An integrative multi-omics approach to quantify biological age with electronic medical records. *Physiology*, 39(S1):581, 2024.

[42] Sheng Fong, Kamil Pabis, Djakim Latumalea, Nomuundari Dugersuren, Maximilian Unfried, Nicholas Tolwinski, Brian Kennedy, and Jan Gruber. Principal component-based clinical aging clocks identify signatures of healthy aging and targets for clinical intervention. *Nature Aging*, 4(8):1137–1152, 2024.

[43] Maria Brbić, Michihiro Yasunaga, Prabhat Agarwal, and Jure Leskovec. Predicting drug outcome of population via clinical knowledge graph. *medRxiv*, 2024.

[44] Shanghua Gao, Richard Zhu, Zhenglun Kong, Ayush Noori, Xiaorui Su, Curtis Ginder, Theodoros Tsiligkaridis, and Marinka Zitnik. TxAgent: An AI agent for therapeutic reasoning across a universe of tools. *arXiv: 2503.10970*, 2025.

[45] Yin Wu, Pierre Beylot, and N Magnenat Thalmann. *Skin aging estimation by facial simulation*. IEEE, 1999.

[46] Wan-Hsuan Lu, Sophie Guyonnet, Jérémy Raffin, Sandrine Bessou-Touya, Katia Ravard Helffer, Pascale Bianchi, Jimmy Le Digabel, Paul Bensadoun, Jean-Marc Lemaitre, Philipe de Souto Barreto, et al. Associations of Skin Biomechanical Properties With Biological Aging Clocks and Longitudinal Changes in Intrinsic Capacity in Adults Aged 20–93: The INSPIRE-T Project. *Aging Cell*, 2025.

[47] Neural Information Processing Systems (NeurIPS) Foundation. NeurIPS Code of Ethics. Online at NeurIPS website: `https://neurips.cc/public/EthicsGuidelines`, 2025. Guidelines on research ethics, broader societal impact, and reviewer responsibilities for NeurIPS submissions.

[48] NeurIPS Conference. NeurIPS2025 Policy on the Use of Large Language Models. NeurIPS official website: `https://neurips.cc/Conferences/2025/LLM`, 2025. Program Chairs: Nancy Chen, Marzyeh Ghassemi, Piotr Koniusz, Razvan Pascanu, Hsuan-Tien Lin; Assistant Program Chairs: Elena Burceanu, Junhao Dong, Zhengyuan Liu, Po-Yi Lu, Isha Puri.

A Technical Appendices and Supplementary Material

A.1 Training (IMDB-FaceTTD)

Data assembly. We use the IMDB–FaceTTD dataset constructed by merging per-image metadata with portrait files and computing time-to-death at photo time (`time_to_death_at_face`). Records with missing `imdb_id` or `time_to_death_at_face` are dropped. Implementation details are in `intrinsic_capacity.ipynb` and `best_clean_icfaceage_notebook.ipynb` (see repository: `https://anonymous.4open.science/r/facettd-CD51/`).

Preprocessing and features. Each portrait is converted to grayscale and resized to 64×64 (PIL `convert("L").resize((64,64))`), then flattened. We concatenate image pixels with numeric metadata: (i) chronological age at photo time; (ii) binary gender (female=1, else 0). In the IMDB pipeline we additionally one-hot encode `cause_of_death` using `OneHotEncoder(handle_unknown="ignore")`. Numeric metadata (age, gender) are standardized with `StandardScaler` fit on the *training* split and applied to validation/test columns in place (feature slice immediately preceding the cause-of-death one-hot block).

Leakage control and splitting. To prevent identity leakage, we split by subject using grouped samplers: (i) *GroupShuffleSplit* with `test_size=0.2, random_state=42`; and (ii) a TTD-stratified subject split via *StratifiedGroupKFold* with n_splits=5, `shuffle=True`, `random_state=42`, using q=10 quantile bins of TTD for stratification. For reporting, we use a single reproducible train/test partition (first SGKF fold), which is approximately 80/20 at the subject level.

Models and hyperparameters. We train two regressors on the pixel-metadata vector: Random Forest (sklearn) and XGBoost (xgboost). When available, we load the best-performing configurations from serialized artifacts (`best_rf.pkl`, `best_xgb.pkl`); otherwise we use the fixed baselines present in the notebooks. The **Random Forest Regressor** is configured with `n_estimators=100`, `max_depth=10`, `n_jobs=-1`, and `random_state=42`. The **XGB Regressor** is configured with `n_estimators=100`, `max_depth=10`, `n_jobs=-1`, `verbosity=0`, and `random_state=42`.

No early stopping or external hyperparameter search is used in the notebooks; the "best" variants are loaded from prior selection and then refit/evaluated as indicated below.

Efficiency. Images are loaded and featurized in parallel using a `ThreadPoolExecutor` (up to 32 workers). All randomization (splitting and model seeds) uses `random_state=42` for reproducibility.

A.2 Evaluation

Metrics and reporting. We report coefficient of determination (R^2) and mean absolute error (MAE). Unless otherwise noted, results are computed once on the held-out test partition induced by the stratified subject split (no multi-seed averaging in the current notebooks). Plots show train vs. test scatter with an identity line.

In-distribution (IMDB-FaceTTD). Training, preprocessing, and model selection are performed on the training portion of the IMDB subject split. The held-out IMDB test subjects are used once for final ID reporting.

Out-of-distribution (Wiki-FaceTTD). We build OOD evaluation data from `wiki.mat` by merging portrait metadata with Wikipedia death years and computing `time_to_death = deathYear - photo_taken` (`00_wiki_extract.ipynb`). The resulting table is then cleaned and preprocessed (`01_wiki_preprocess.ipynb`): portraits are converted to grayscale, resized to 64×64, and flattened; photo age is standardized; and entries missing valid death years or corrupted images are removed. The notebook outputs aligned feature arrays that match the IMDB model expectations.

At test time (`02_wiki_test.ipynb`), these preprocessed features are loaded and concatenated with the standardized age at photo time. When model input dimensionality exceeds available wiki features (e.g., models trained with IMDB-specific one-hot cause-of-death and/or gender), we align dimensions by zero-padding or truncation to the model's `n_features_in_`; gender is set to missing for wiki (`NaN`) and not used for scaling. We evaluate pre-trained IMDB models on wiki *without refitting* and report R^2 and MAE overall and under clinically motivated TTD filters: TTD ≤ 60 years, TTD ≤ 45 years, and $5 \leq$ TTD ≤ 45 years.

Longitudinal coverage analysis (IMDB). To probe temporal coverage, we annotate each subject by the span and count of available portraits: *Point* (single portrait), *1-Year* ($>$1 portrait with age span $\leq$1 year), *Full* ($>$1 portrait with span $>$1 year), and *Year+Full* (any subject with $>$1 portrait). Per-condition train/test subsets are constructed at the subject level, and both Random Forest and XGBoost are evaluated within each condition (`best_clean_icfaceage_notebook.ipynb`).

Reproducibility. All ID/OOD code paths are referenced in the notebooks: `intrinsic_capacity.ipynb`, `best_clean_icfaceage_notebook.ipynb` (IMDB); `00_wiki_extract.ipynb`, `01_wiki_preprocess.ipynb`, `02_wiki_test.ipynb` (wiki). The full repository, including preprocessing and evaluation code, is available at `https://anonymous.4open.science/r/facettd-CD51/`. Random seeds are fixed at 42 for splits and model initialization. Scalers used at test time (e.g., age standardization) are loaded from the saved training artifacts.

A.3 SHAP Analysis of Age-Dependent and Age-Independent Signals

To interpret the model's predictions, we computed SHAP (SHapley Additive exPlanations) values, which decompose the model output as $f(\mathbf{x}) = \phi_0 + \sum_i \phi_i$, where ϕ_i denotes the marginal contribution of feature i relative to the model's expected output ϕ_0.

The SHAP analysis revealed that the model's predictive structure remains consistent with or without chronological age as an input. When age was included, it contributed to baseline variance but did not alter the relative influence or directionality of other features, indicating that the model captures intrinsic phenotypic patterns beyond explicit age information. When age was excluded, overall performance remained comparable, suggesting that the learned representations encode latent aging signals independent of chronological age.

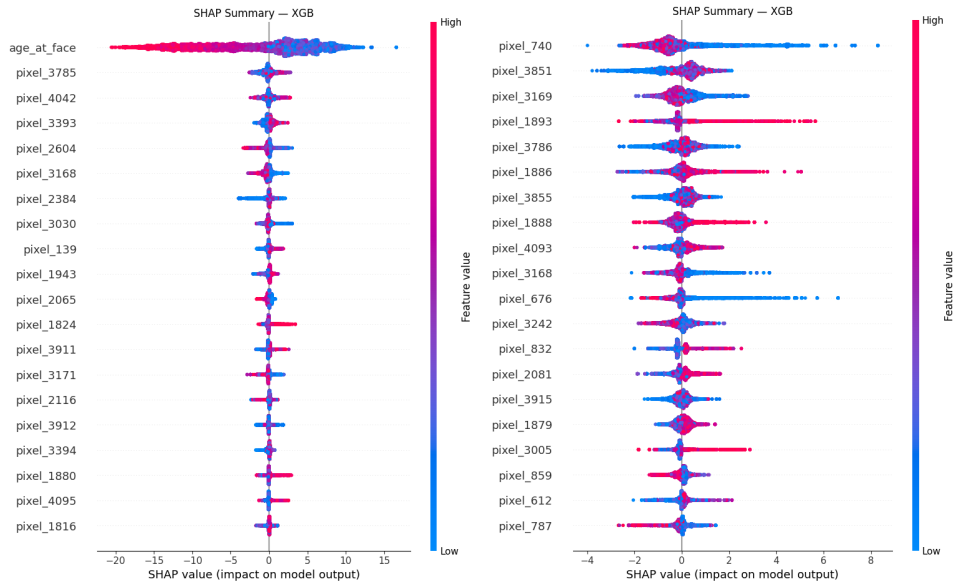

Figure 5: SHAP summary plots for the XGBoost model trained (left) with and (right) without chronological age as an input feature. The age signal dominates the model when included, yet the persistence of pixel-level contributions in the age-excluded model suggests that facial features independently encode predictive aging information.

NeurIPS Paper Checklist

1. **Claims** Answer: [Yes] . Justification: The abstract and introduction state that the paper develops ICFaceAge, predicts time-to-death from facial time series, reports in- and out-of-distribution results, and these claims are supported by experiments.
2. **Limitations** Answer: [Yes] . Justification: Section 4.1 Limitations outlines dataset bias, reliance on reported death dates, and model constraints.
3. **Theory assumptions and proofs** Answer: [NA] . Justification: No new theoretical results are introduced; the work is empirical.
4. **Experimental result reproducibility** Answer: [Yes] . Justification: Data sources, preprocessing steps, model choices, and evaluation methods are described; a HuggingFace demo is provided.
5. **Open access to data and code** Answer: [Yes] . Justification: Demo code and curated scripts are available via HuggingFace.
6. **Experimental setting/details** Answer: [Yes] . Justification: Data splits, regressors, and evaluation metrics are specified; hyperparameter tuning described.
7. **Experiment statistical significance** Answer: [No] . Justification: Error bars are not reported due to computational cost; results are presented with R^2 and MAE only.
8. **Experiments compute resources** Answer: [Yes] . Justification: Training was performed on a workstation with standard CPU/GPU; requirements are minimal.
9. **Code of ethics** Answer: [Yes] . Justification: Research conforms to NeurIPS Code of Ethics, with explicit acknowledgment of societal risks.
10. **Broader impacts** Answer: [Yes] . Justification: Positive and negative societal impacts (e.g., health applications vs surveillance misuse) are discussed in Discussion.
11. **Safeguards** Answer: [Yes] . Justification: A usage disclaimer is included on the HuggingFace public demo as a safeguard, stating that it is intended solely for research purposes.
12. **Licenses for existing assets** Answer: [Yes] . Justification: IMDB and Wikipedia are publicly available sources; terms of service respected.
13. **New assets** Answer: [Yes] . Justification: New curated datasets and preprocessing steps are documented with the released code in the repository.
14. **Crowdsourcing and research with human subjects** Answer: [NA] . Justification: Work uses publicly available images; no crowdsourcing or direct human subjects research.
15. **Institutional review board (IRB) approvals or equivalent** Answer: [NA] . Justification: No direct human subjects research conducted; data from public sources only.
16. **Declaration of LLM usage** Answer: [Yes] . Justification: Large language models (LLMs) were used broadly in the preparation of this work, including for text drafting and editing, code generation, and figure preparation. All methodological design, experimental execution, data analysis, and scientific conclusions were conceived and validated by the authors. LLM usage did not substitute for scientific reasoning and does not affect the rigor, validity, or integrity of the work.