Predicting Biological Age of ICU Patients from MIMIC-IV Using TabM

ELLIOT FARMER GARCIA, University of Texas at Austin, USA

Estimating a patient's biological age from routinely collected ICU data offers a window into physiologic reserve and potential trajectories in critical care. We curate a cohort of 72 338 ICU admissions from MIMIC-IV, extracting 33 features—including demographics (gender, insurance, language, marital status, race), length of stay, and admission-averaged vital signs and laboratory measurements. Missing categorical values are imputed using the most frequent category and encoded ordinally, while continuous variables are imputed via KNN ($k = 5$) and standardized. We compare a simple multilayer perceptron (MLP) baseline against TabM, a lightweight BatchEnsemble MLP, both with and without categorical inputs. On a held-out validation set, the MLP baseline achieves an RMSE of 11.28 years, TabM without categorical inputs yields an RMSE of 11.04 years, and TabM with categorical inputs reduces RMSE to 9.24 years—an $\approx 18\%$ reduction relative to the baseline. These results demonstrate that TabM, when paired with principled imputation and categorical encoding, yields a promising biological-age predictor in the ICU.

CCS Concepts: • **Computing methodologies** → **Neural networks**; **Artificial intelligence**; • **Applied computing** → **Health informatics**.

Additional Key Words and Phrases: biological age prediction, electronic health records, ICU, tabular deep learning, MIMIC-IV

ACM Reference Format:

Elliot Farmer Garcia. 2025. Predicting Biological Age of ICU Patients from MIMIC-IV Using TabM. 1, 1 (April 2025), 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Accurate assessment of a patient's physiological state is a cornerstone of effective critical-care management. Chronological age is routinely used as a crude proxy for biological resilience, yet it fails to capture the substantial heterogeneity in how individuals age and respond to stressors. Two patients of the same calendar age may present with vastly different organ reserves and comorbidity burdens, leading to under- or overestimation of their risk during an intensive-care unit (ICU) stay. A reliable "biological clock" that quantifies physiological age from routinely collected electronic health record (EHR) data could enable more personalized triage, prognostication, and resource allocation in critical care.

In this work, we address the challenge of estimating chronological age from multimodal ICU data—demographics, laboratory tests, and vital-sign trends—as a proxy for true biological age. We leverage the publicly available MIMIC-IV dataset, encompassing over 70 000 ICU admissions, and apply a state-of-the-art tabular deep-learning architecture (TabM) alongside a simple multilayer perceptron baseline. Our goal is twofold: first, to demonstrate that modern ensemble-inspired MLPs can learn a robust mapping from high-dimensional EHR features to age with significantly lower error than traditional baselines; and second, to establish a scalable pipeline—including KNN imputation and standardization—that can serve as a blueprint for future clinical-age-prediction models.

Author's Contact Information: Elliot Farmer Garcia, efgarcia@utexas.edu, University of Texas at Austin, Austin, Texas, USA.

Manuscript submitted to ACM

By improving age-prediction accuracy from an RMSE of 11.3 years with a standard MLP to 9.2 years with TabM, we show meaningful progress toward a practical biological clock for ICU settings—one that could inform dynamic patient monitoring, risk stratification, and ultimately, more tailored therapeutic decisions.

2 Related Work

Prior efforts in biological age prediction have leveraged diverse data sources and modeling strategies.

Levine et al. [3] introduced DNAm PhenoAge, an epigenetic clock trained via penalized proportional-hazards models on nine clinical biomarkers and chronological age, demonstrating improved mortality and morbidity prediction over chronological age alone.

Sagers et al. [4] developed a supervised machine-learning approach using 356 blood laboratory measurements from 67,563 individuals, achieving a mean absolute error of 4.76 years in held-out data and highlighting the value of large-scale lab biomarkers for age estimation.

More recently, Chen et al. [1] proposed OMICmAge, an integrative multi-omics framework that combines genomic, proteomic, and electronic health record data to produce a comprehensive biological age metric validated across biobank cohorts.

While these studies illustrate the power of high-dimensional biomarkers and advanced modeling, they generally focus on broad population cohorts and multi-omics inputs. In contrast, our work concentrates on ICU admissions in MIMIC-IV and demonstrates that a lightweight BatchEnsemble MLP (TabM), applied to routinely collected vitals and laboratory averages with principled imputation, can yield competitive age-prediction performance in a critical-care setting.

3 Methods

3.1 Data Source and Cohort Selection

We used the publicly available MIMIC-IV database and restricted our analysis to ICU admissions in order to utilize chart events. After joining `admissions`, `icustays`, `labevents`, and `chartevents`, our final cohort comprised 72 338 ICU stays.

3.2 Feature Extraction

For each admission, we extracted:

- **Demographics:** gender, insurance, language, marital status, race.
- **Length of stay:** time between `admittime` and `dischtime`, in days.
- **Laboratory values:** admission-average of 20 most frequently ordered blood tests (e.g. glucose, creatinine, electrolytes).
- **Vital signs:** admission-average of heart rate, non-invasive blood pressure (systolic/diastolic), respiratory rate, SpO_2, and temperature.
- **Age:** chronological age at admission, calculated by aligning the deidentified admission year to the patient's `anchor_year` and adding the difference to the `anchor_age`.

3.3 Data Preprocessing

- **Categorical features** (demographics): imputed with the most frequent category using `SimpleImputer(strategy="most_frequent")`, then encoded ordinally.
- **Continuous features** (labs, vitals, length of stay): imputed using `KNNImputer(n_neighbors=5)`, then standardized to zero mean and unit variance.
- **Target feature** (age): standardized to zero mean and unit variance.

3.4 Model Architectures

We evaluated two architectures:

(1) **MLP baseline:** a simple network with two blocks of fully-connected layers with ReLU activations, layer normalization, and 0.3 dropout.
(2) **TabM:** a parameter-efficient ensembled MLP [2] producing $k = 32$ implicit submodels that share the majority of weights via BatchEnsemble adapters.

3.5 Training and Evaluation

We performed an 80/20 split into training and validation sets. All preprocessing (imputation + scaling) was fit on the training fold and applied to both splits in order to prevent data leakage. All sources of randomness were seeded for deterministic output. Models were trained for 100 epochs with:

- **Batch size:** 256
- **Optimizer:** AdamW with learning rate 2×10^{-3} and weight decay 3×10^{-4}
- **Loss:** mean squared error (MSE) on standardized age
- **Scheduler:** ReduceLROnPlateau (factor=0.5, patience=10) on validation loss

Final performance was reported as RMSE on the original age scale (years) by unscaling MSE via multiplication by σ_y^2 and taking the square root.

Figure 1 illustrates the end-to-end pipeline from raw data extraction through model evaluation.

4 Results

Our experiments demonstrate that the proposed TabM architecture significantly improves prediction accuracy over a simple multilayer perceptron (MLP) baseline. All errors are reported as unscaled root–mean–square error (RMSE) in years, obtained by back-transforming the model outputs using the target population mean (62.67 years) and standard deviation (16.18 years).

4.1 Predictive Performance

Model	RMSE (years)	Improvement vs. MLP
MLP (continuous only)	11.28	—
TabM (continuous only)	11.04	2.1%
TabM (continuous + categorical)	9.24	18.1%

Table 1. Comparison of unscaled RMSE for age prediction models.

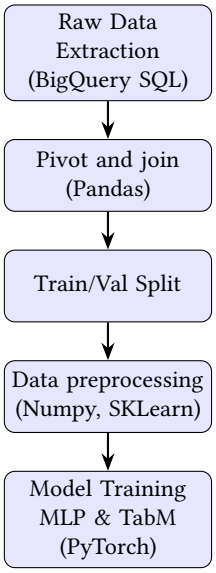

Fig. 1. End-to-end biological-age pipeline: from raw MIMIC-IV extraction to model evaluation.

4.2 Variance Explained

By comparing unscaled MSE to the total variance of chronological age ($\sigma^2 = 16.1793^2 \approx 261.8$):

$$R^2 \;=\; 1 - \frac{(9.23)^2}{261.8} \;\approx\; 0.67,$$

indicating that our best model explains approximately 67% of the variation in patient age.

4.3 Summary of Results

These results demonstrate that parameter-efficient ensembling (TabM) yields modest gains over a basic MLP when using only continuous laboratory and chart features, and that incorporating key categorical demographics further boosts accuracy by nearly two years of RMSE. This level of precision ($\approx$ 9.2 years) represents a substantial advance toward practical biological-age estimation from routine clinical data.

5 Conclusion

In this work, we developed a lightweight biological-age predictor using routinely collected laboratory and demographic data from MIMIC-IV. By comparing a simple multilayer perceptron (MLP) baseline to our parameter-efficient TabM model and its extension incorporating categorical covariates, we demonstrated progressive improvements in validation RMSE—from 11.28 years (MLP) to 11.04 years (TabM) and finally to 9.24 years (TabM + categoricals). These results illustrate that efficient ensembling of MLPs can yield highly accurate age estimates without resorting to expensive or high-dimensional omics data.

Future work can explore several avenues for enhancing and extending this approach:

- **Interpretability and feature importance:** Apply model-agnostic methods (e.g., SHAP) to identify the most informative biomarkers and improve clinical trust.

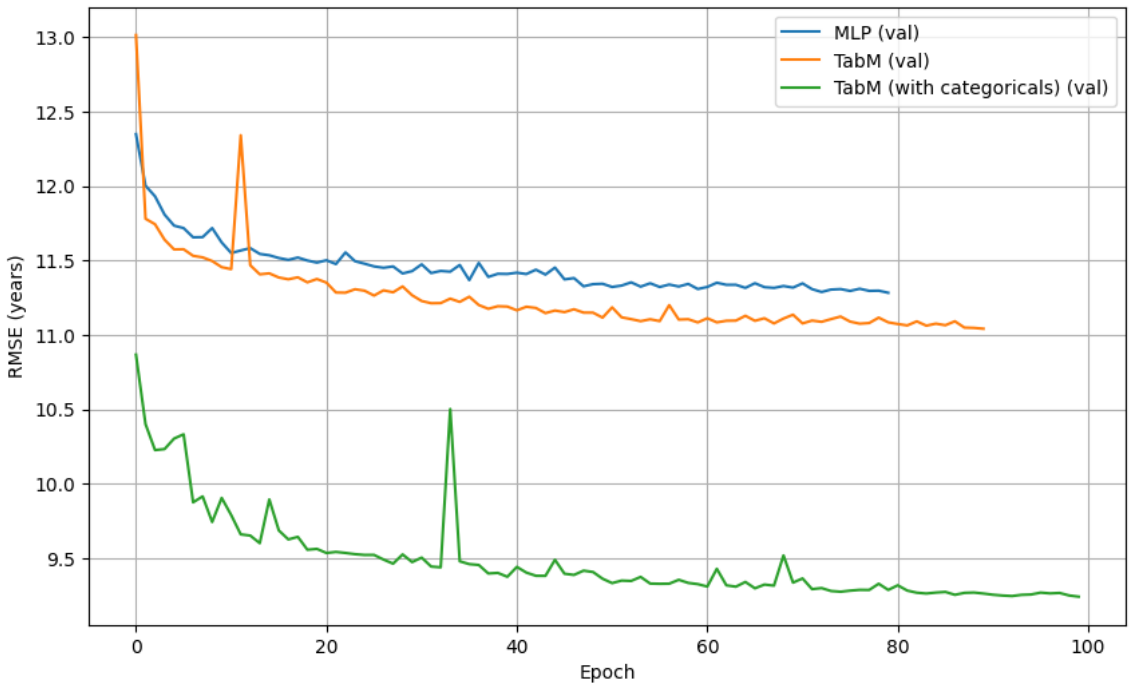

Fig. 2. Model loss over training epochs on the validation set.

- **Uncertainty quantification:** Leverage the ensemble nature of TabM to derive prediction intervals and support decision-making in individual patients.
- **Temporal modeling:** Extend the pipeline to predict *biological age acceleration* and its longitudinal trajectories using recurrent or attention-based architectures.
- **External validation:** Test the model on other hospital systems or population cohorts to assess generalizability and robustness.

Overall, our findings highlight the promise of efficient deep-learning ensembles for scalable, cost-effective biological-age estimation and lay the groundwork for richer, clinically actionable models in the future.

References

[1] Qingwen Chen, Varun B. Dwaraka, Natàlia Carreras-Gallo, Kevin Mendez, Yulu Chen, Sofina Begum, Priyadarshini Kachroo, Nicole Prince, Hannah Went, Tavis Mendez, Aaron Lin, Logan Turner, Mahdi Moqri, Su H. Chu, Rachel S. Kelly, Scott T. Weiss, Nicholas J.W Rattray, Vadim N. Gladyshev, Elizabeth Karlson, Craig Wheelock, Ewy A. Mathé, Amber Dahlin, Michae J. McGeachie, Ryan Smith, and Jessica A. Lasky-Su. 2023. OMICmAge: An integrative multi-omics approach to quantify biological age with electronic medical records. *bioRxiv* (2023). doi:10.1101/2023.10.16.562114

[2] Yury Gorishniy, Akim Kotelnikov, and Artem Babenko. 2025. TabM: Advancing Tabular Deep Learning with Parameter-Efficient Ensembling. arXiv:2410.24210 [cs.LG]

[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. 2018. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* 10, 4 (April 2018), 573–591. doi:10.18632/aging.101414

[4] Luke Sagers, Luke Melas-Kyriazi, Chirag J. Patel, and Arjun K. Manrai. 2020. Prediction of chronological and biological age from laboratory data. *Aging (Albany NY)* 12, 9 (May 2020), 7626–7638. doi:10.18632/aging.102900