Predicting Alzheimer's Disease Progression Using a Regression-based Survival Model with Longitudinal Data

Ling Dai[1], Yiqun Sun[1,2], Jincheng Gu[1], Qinsen Bao[3], Feihong Liu[1], Dinggang Shen[1,4,5]

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials & Devices, ShanghaiTech University, Shanghai 201210, China
[2] Department of Radiology, Fudan University Shanghai Cancer Center, Shanghai, China
[3] School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
[4] Shanghai United Imaging Intelligence Co. Ltd., Shanghai 200230, China
[5] Shanghai Clinical Research and Trial Center, Shanghai 201210, China
dgshen@shanghaitech.edu.cn

Abstract. Alzheimer's disease (AD) progression is characterized by slow, heterogeneous, and subtle changes that span decades, making transition points difficult to determine. This challenge is compounded by the complexity of longitudinal clinical data, including irregular follow-up patterns and varying observation durations that traditional survival analysis models cannot handle. We present a novel regression-based survival framework with three key innovations: (1) Robust longitudinal data handling, (2) Enhanced early-stage prediction, and (3) Flexible integration with existing models. Using a partial optimization approach for mean squared error loss, our method achieves state-of-the-art performance in AD progression prediction, particularly excelling in early-stage scenarios. Ablation studies identify the regression loss component as the key driver of improved long-term prediction capability, advancing AD prognosis and broader applications in longitudinal survival analysis.

Keywords: Alzheimer's disease · Survival analysis · Machine learning · Longitudinal data

1 Introduction

Alzheimer's disease (AD), the leading cause of dementia and an irreversible, progressive neurodegenerative disorder, makes early diagnosis essential for timely intervention to slow its pathological progression [1]. For early AD diagnosis, structural neuroimaging has recently been widely employed owing to its non-invasive nature, its ability to provide high-resolution images of critical brain structures (such as the medial temporal and hippocampal regions), and its capacity for quantitative atrophy analysis that yields objective biomarkers [2].

Moreover, recent progress in deep learning, especially with architectures like generative models and foundation models, excels at capturing complex relationships and long-range dependencies within image data, and significantly improves diagnostic accuracy [3, 4]. However, precise diagnosis may fail when dealing with subtle and heterogeneous AD progression [5].

Survival analysis models have been developed to model pathological progression. Nevertheless, conventional methods may fail due to incomplete progression time caused by censoring [6, 7]. Recent advances address this issue by estimating the survival function $S(t)$, which serves as a probabilistic model to predict the likelihood that an individual will remain disease-free beyond a specific time point t [8, 9]. More specifically, DeepHit [10], as a representative model, is designed to divide the timeline into fixed-length intervals, which estimates the probability mass function for each interval and accumulates these probabilities to construct the survival function, $S(t)$ [11]. Another approach extends the Cox proportional hazards model by parameterizing the hazard function using neural networks as exemplified by DeepSurv [12]. A third approach directly estimates $S(t)$ through a distributional assumption on the progression time parameterized by a deep learning model as described in Centime [13]. Unlike binary classification, this approach models the likelihood of remaining disease-free beyond time t, offering deeper temporal insights and enabling earlier detection of subtle changes in disease evolution.

In the context of AD, its pathological progression is slow, subtle, and heterogeneous, making determining the time-to-progression (TTP) inherently challenging, even in datasets without censoring. Even worse, real-world data often feature irregular visit patterns and diverse follow-up spans. On the one hand, stringent data requirements restrict the amount of training data available to the models, leading to poorer performance. On the other hand, the complex distribution of real-world data violates the models' prior assumptions, introducing bias during the training process.

In this paper, we address these challenges using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [14]. Our method directly estimates the TTP from longitudinal datasets and offers several key advantages over existing survival models: (1) *Enhanced longitudinal data utilization*: Unlike conventional survival models that primarily focus on survival function estimation, our approach explicitly models intra-subject temporal differences to capture disease progression patterns; (2) *Improved early-stage prediction*: By leveraging pairwise temporal relationships within subjects, our framework can extrapolate disease trajectories even with sparse initial observations; (3) *Flexible model integration*: Our regression-based approach can be seamlessly integrated with established survival models as a regularization term, enhancing their long-term predictive accuracy while preserving model-specific characteristics. Our approach leverages the full richness of longitudinal data and explicitly models disease dynamics, enabling precise, individualized predictions of AD progression.

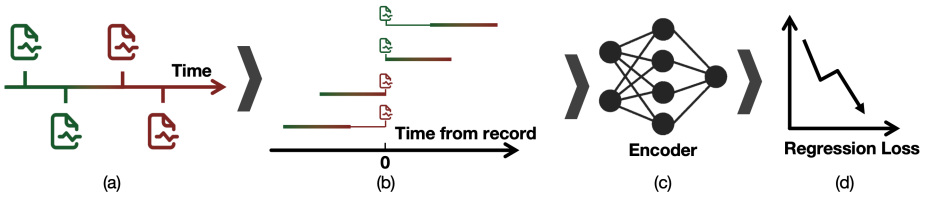

Fig. 1. Illustration of the proposed regression-based survival model framework. (a) Longitudinal records of a subject with green/red markers indicating pre-/post-AD transition events ($s_i = 0/1$), (b) Temporal bound transformation mapping each observation to its transition time region $[t^-, t^+]$, where green endpoints represent lower bounds t_i^- and red endpoints denote upper bounds t_i^+, (c) Neural encoder architecture with MLP backbone producing time-to-progression predictions $y = f_\theta(\mathbf{X})$, and (d) Regression loss components L_{reg} (pairwise temporal difference matching) and L_{bias} (censoring-aware likelihood). Arrows indicate information propagation through the survival prediction pipeline.

2 Proposed Methods

2.1 Preliminaries

Let $\mathcal{S}_{\text{long}}$ denote the set of longitudinal records consisting of N observations across M subjects, where each record is represented as a tuple $\langle \mathbf{x}_i, t_i, o_i, e_i \rangle$. In this context, $\mathbf{x}_i$ denotes neuroimaging and clinical features observed at time t_i, $o_i \in \{1, 2, \ldots, M\}$ is a unique identifier for the subject, and $e_i \in \{0, 1\}$ indicates the event status (0: no event; 1: event occurred).

Our objective is to develop a parameterized regression model $f_\theta(\mathbf{x})$ for estimating T, where T denotes the time-to-progression from non-AD to AD. For each record i, we define temporal regions t_i^- and t_i^+, representing the relative time to the last visit before progression and the first visit after progression, respectively. For right-censored data (where progression has not occurred), we set $t_i^+ = \max_{j:o_j=o_i} t_j - t_i + C$. Similarly for left-censored data (where progression occurs before the first visit), we set $t_i^- = t_i - \min_{j:o_j=o_i} t_j - C$. Here, C is a sufficiently large constant, effectively treated as infinity. This definition facilitates derivations when both records are from the same subject, $t_i^+ - t_j^+ = t_i^- - t_j^- = T_i - T_j$. Each record can be represented as $\langle \mathbf{x}_i, o_i, t_i^-, t_i^+ \rangle$ since t_i^- and t_i^+ encapsulate all necessary survival information.

2.2 Regression Loss for Censored Time-to-Progression Prediction

While minimizing mean-squared error (MSE) is ideal for training a regression model to predict TTP, censoring precludes direct observation of the target variable T. We tackle this through partial optimization of the MSE loss function. Let $y_i = f_\theta(\mathbf{x}_i)$ denote the predicted TTP and $d_i = T_i - y_i$ represent the residual error. The mean residual error is computed as $\bar{d} = \frac{1}{N} \sum_{i=1}^{N} d_i$. The MSE can be

decomposed as follows:

$$L_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}\left(d_i^2\right) = \frac{1}{N}\sum_{i=1}^{N}\left(d_i - \bar{d} + \bar{d}\right)^2 \tag{1}$$

$$= \frac{2}{N}\sum_{i=1}^{N}\bar{d}\left(d_i - \bar{d}\right) + \frac{1}{N}\sum_{i=1}^{N}\left(d_i - \bar{d}\right)^2 + \frac{1}{N}\sum_{i=1}^{N}\bar{d}^2. \tag{2}$$

Equation 2 consists of three additive components; we will examine each component to eliminate all latent variables T_i and d_i.

First Term of Equation 2

$$\sum_{i=1}^{N}\bar{d}\left(d_i - \bar{d}\right) = \bar{d}\sum_{i=1}^{N}\left(d_i - \frac{1}{N}\sum_{j=1}^{N}d_j\right) = 0.$$

Second Term of Equation 2

$$\frac{1}{N}\sum_{i=1}^{N}\left(d_i - \bar{d}\right)^2 \leq \frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=1}^{N}\left(d_i - d_j\right)^2 = \frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=1}^{N}\left(T_i - T_j - (y_i - y_j)\right)^2.$$

Note that $T_i - T_j$ is available only if both records originate from the same subject, and $T_i - T_j = t_i^+ - t_j^+$ as described in Section 2.1. Furthermore, if neither o_i nor o_j is censored, and assuming that t_i^+ and t_j^+ follow unknown distributions with probability density functions $g_i(t)$ and $g_j(t)$, respectively, we have:

$$\begin{aligned} E(t_i^+ - t_j^+) &= \iint (t_i^+ - t_j^+) g_i(t_i^+) g_j(t_j^+)\, dt_i^+ dt_j^+ \\ &= \iint t_i^+ g_i(t_i^+) g_j(t_j^+)\, dt_j^+ dt_i^+ - \iint t_j^+ g_i(t_i^+) g_j(t_j^+)\, dt_i^+ dt_j^+ \\ &= \mathbb{E}(T_i - T_j). \end{aligned}$$

This demonstrates that substituting $T_i - T_j$ with $t_i^+ - t_j^+$ does not introduce additional bias. Therefore, we can define L_{reg} as:

$$L_{\text{reg}} = \sum_{1 \leq i \leq j \leq N, o_i = o_j \vee -C < t_i^+, t_j^+, t_i^-, t_j^- < C} \left(t_i^+ - t_j^+ - (y_i - y_j)\right)^2. \tag{3}$$

The design of L_{reg} effectively captures *intra-subject variations* between records, providing additional temporal information for the model in handling censored data. Additionally, the differencing mechanism successfully addresses *inter-subject heterogeneity* without introducing bias when T is latent.

Third Term of Equation 2

According to the definition of d, $\frac{1}{N}\sum_{i=1}^{N}\bar{d}^2 = \bar{d}^2 = \left(\frac{1}{N}\sum_{i=1}^{N}(T_i - y_i)\right)^2$. Since T_i remains unknown, we employ likelihood maximization to minimize this expression. Assuming that T follows a distribution with mean μ, Bayesian inference yields:

$$\mathbb{P}\left(\mu = y_i \mid t_i^+, t_i^-\right) = \frac{\mathbb{P}\left(t_i^+, t_i^- \mid \mu\right)\mathbb{P}\left(\mu = y_i\right)}{\mathbb{P}\left(t_i^+, t_i^-\right)} \propto P_{y_i}\left(t_i^+\right)\left(1 - P_{y_i}\left(t_i^-\right)\right)p\left(y_i\right), \tag{4}$$

where $P_y(t)$ and $p(y)$ denote the cumulative distribution function (CDF) of $\mathcal{D}_\mu^T$ and the probability density function (PDF) of $\mathcal{D}^y$, respectively. We model T using a logistic distribution with mean μ and scale s, such that $P_\mu(t) = \frac{1}{1+\exp(-(t-\mu)/s)}$, and assume a normal prior $\mathcal{N}(0, \sigma^2)$ for y. The negative log-likelihood loss L_{bias} is defined as:

$$L_{\text{bias}} = -\sum_{i=1}^{N}\log\frac{1}{1+e^{-(t_i^+ - y_i)/s}}\left(1 - \frac{1}{1+e^{-(t_i^- - y_i)/s}}\right)e^{-y_i^2/\sigma^2}. \tag{5}$$

The L_{bias} loss function is designed to estimate latent TTP across potential regions in the absence of accurate TTP.

Full Regression Loss Combining Equation 5 and Equation 3, we define the regression loss for censored TTP prediction:

$$L_{\text{Regression}} = L_{\text{reg}} + \alpha L_{\text{bias}}. \tag{6}$$

where α is a hyperparameter used to control the strength of the bias loss.

2.3 Model Optimization and Integration with Existing Methods

The proposed differentiable loss function enables seamless integration with most encoder architectures that yield continuous output and facilitates optimization via gradient-based methods. Our regression framework directly estimates TTP, ensuring compatibility with established survival analysis algorithms that infer expected TTP, denoted as output y in Section 2.2. Integration is achieved by incorporating the regression loss as regularization within existing survival analysis algorithms. Given an original objective function L_{original}, the combined loss is defined as $L_{\text{Total}} = L_{\text{Original}} + \lambda L_{\text{Regression}}$, where λ governs regularization strength. This integration is theoretically justified because existing survival models estimate the survival function $S(t)$, whereas our regression framework directly predicts TTP. By combining these losses, we align the predicted TTP with the expectation of the survival distribution, ensuring consistency between the two objectives while leveraging the complementary strengths of both approaches. This minimally invasive integration preserves the characteristics of baseline models while enhancing performance, as demonstrated experimentally.

Table 1. Performance comparison of AD progression prediction models using C-index and mean AUC (with 95% confidence intervals). * denotes our proposed model (Regression) or existing models enhanced by our framework (Regression+[Base Model]), with Regression+DeepHit achieving the best results (bold).

Model	C-index (95% CI)	Mean AUC (95% CI)
Centime [13]	0.714 (0.660-0.746)	0.700 (0.654-0.746)
DeepHit [10]	0.743 (0.715-0.770)	0.775 (0.742-0.814)
DeepSurv [12]	0.676 (0.639-0.703)	0.684 (0.631-0.713)
*Regression	0.765 (0.727-0.797)	0.794 (0.761-0.821)
*Regression+Centime	0.734 (0.701-0.767)	0.753 (0.720-0.782)
*Regression+DeepHit	**0.782 (0.749-0.811)**	**0.804 (0.772-0.830)**
*Regression+DeepSurv	0.747 (0.719-0.776)	0.772 (0.738-0.799)

3 Experiments and Results

Dataset Description Our study utilized structural magnetic resonance imaging (sMRI) data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), comprising 1,754 subjects (50.22% male, mean age 73.5 ± 7.14 years) with 6,400 longitudinal records. The cohort was followed for 2.66 ± 2.95 years (maximum 14.07 years), during which 252 subjects progressed to AD. The large standard deviation in follow-up duration reflects the long-tailed distribution characteristics of the dataset, where a few subjects had very long tracking periods (10–14 years), while most subjects had shorter follow-up periods. All sMRI data underwent standardized preprocessing (N4 bias field correction [15], skull-stripping, ACPC alignment [16]), followed by ROI extraction using the MUSE framework [17] and ComBat harmonization [18] to minimize acquisition variations. The volumetric measurements of 144 anatomically defined brain regions were extracted, which, combined with demographic features (age and gender), yielded a 146-dimensional feature vector for AD progression prediction. To prevent data leakage, the dataset was split at the subject level in a 6:4 ratio for training and testing, ensuring that all records from the same subject remained within the same split. Additionally, 20% of the training data was reserved for parameter tuning, and all models were evaluated using 5-fold cross-validation.

Model Architecture and Experimental Setting We evaluated our proposed regression-based survival model against three state-of-the-art baseline methods, including Centime [13], DeepHit [10], and DeepSurv [12]. For fair comparison, we used a 4-layer MLP encoder with 146 input dimensions and two hidden layers (64 and 16 neurons) for all models. The output layer dimension varies depending on the model requirements: 1 for regression and DeepSurv, 2 for Centime, and 16 for DeepHit. ReLU activation and dropout regularization were implemented between layers. All hyperparameters, including learning rate, weight decay rates, loss component weights (particularly the α parameter controlling the strength of L_{bias}), and regularization strength λ, were optimized

using Bayesian optimization via the *skopt* package [19]. Models were trained using the Adam optimizer with early stopping to prevent overfitting. Final results were obtained through model ensembling from 5-fold cross-validation to ensure robust performance evaluation.

Performance Comparison We used the concordance index (C-index) [20] and averaged time-dependent-AUC [21] as the performance metrics, with results presented in Table 1. Our standalone regression model demonstrated strong performance, achieving a C-index of 0.765 (95% CI: 0.727–0.797) and a mean AUC of 0.794 (95% CI: 0.761–0.821), substantially outperforming traditional approaches. Integration with existing methods via the Regression+DeepHit framework further enhanced performance, yielding a C-index of 0.782 (95% CI: 0.749–0.811) and a mean AUC of 0.804 (95% CI: 0.772–0.830). This represents a significant improvement over the baseline DeepHit model (C-index: 0.743, mean AUC: 0.775), underscoring the effectiveness of our regression-based enhancement. The superior performance of Regression+DeepHit compared to integration with Centime or DeepSurv can be attributed to the compatibility between DeepHit's ranking loss and our regression framework. Both approaches model pairwise relationships between observations, with DeepHit's ranking loss sharing similarities with our L_{reg} component. In contrast, Centime and DeepSurv rely on parametric assumptions (distributional and proportional hazards, respectively) that may conflict with our regression-based approach. To stratify patients into high-/low-risk groups, we use the mean TTP from the training set as a threshold, classifying subjects with predicted TTP exceeding the threshold as low-risk and others as high-risk. The Kaplan-Meier curves in Fig. 2(a) illustrate our model's ability to effectively differentiate risk groups, as evidenced by distinct trajectories for high-risk and low-risk groups.

Predictive Power of Regression Model at Early Stages Our regression-based approach demonstrates superior predictive capability in early disease stages compared to conventional methods, particularly over longer time intervals post-enrollment. This enhanced performance is evident from Fig. 2(b), showing consistently higher time-dependent AUC values as the prediction window extends, offering a valuable clinical window for therapeutic intervention. The improved

Table 2. Ablation study shows the contributions of different loss components: L_{bias} and L_{reg} for the Regression model, and additional L_{rank} from DeepHit for the integrated Regression+DeepHit model.

Regression			**Regression+DeepHit**		
Component	**C-Index**	**Mean AUC**	**Component**	**C-Index**	**Mean AUC**
L_{bias}	0.643	0.660	L_{bias}	0.724	0.749
L_{reg}	0.758	0.783	L_{rank}	0.743	0.775
Full Model	0.765	0.794	L_{reg}	0.780	0.799
-	-	-	Full Model	0.782	0.804

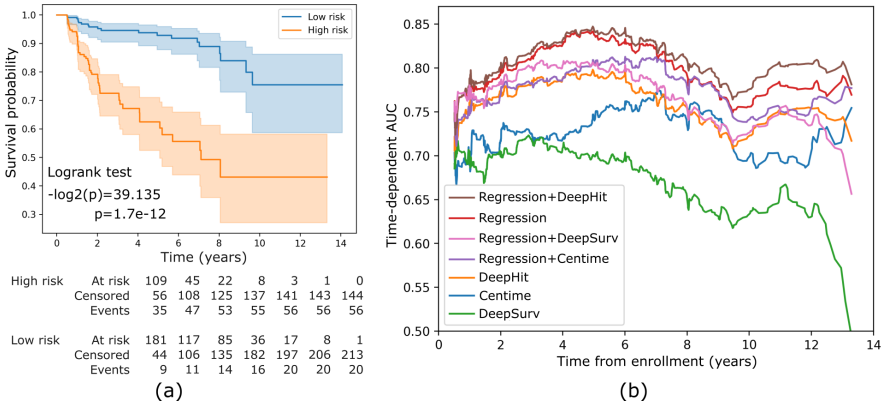

Fig. 2. Performance evaluation of survival prediction models. (a) Kaplan-Meier survival curves show significant stratification between high-risk and low-risk groups as predicted by our best-performing model. (b) Time-dependent AUC comparison across different models demonstrates superior performance of the regression-based approaches over traditional methods.

early-stage performance stems from our model's ability to generalize patterns from limited early observations by modeling pairwise temporal differences across visits (via L_{reg}). This approach enables the framework to extrapolate disease trajectories even with sparse initial data, leveraging the smoothness assumption in neurodegenerative progression where subjects with similar early biomarker profiles and visit patterns are inferred to have comparable time-to-progression. Ablation studies (Table 2) reveal that L_{reg} significantly outperforms L_{bias} in predictions, especially in censored test cases. Analysis of DeepHit's ranking loss (L_{rank}) shows similarities to our L_{reg}, with models utilizing only L_{reg} or L_{rank} demonstrating superior performance compared to those using L_{bias}. This underscores the effectiveness of L_{reg} in enhancing predictive accuracy.

Model Interpretability We use SHAP [22] to interpret the proposed regression model. Results are shown in Fig. 3. Nine features with contributions greater than 0.1 are displayed. The top 3 features are from Left Hippocampus Volume, Right Amygdala Volume, and Left Inf Lat Vent (inferior lateral ventricle), which are consistent with previously identified biomarkers [23, 24]. The explanation reveals how the model predicts the TTP, providing insights into neurodegenerative disease research.

4 Conclusion

In this paper, we present a novel regression-based survival analysis framework that addresses critical challenges in longitudinal medical data analysis by effectively handling irregular temporal sampling patterns and censored observations

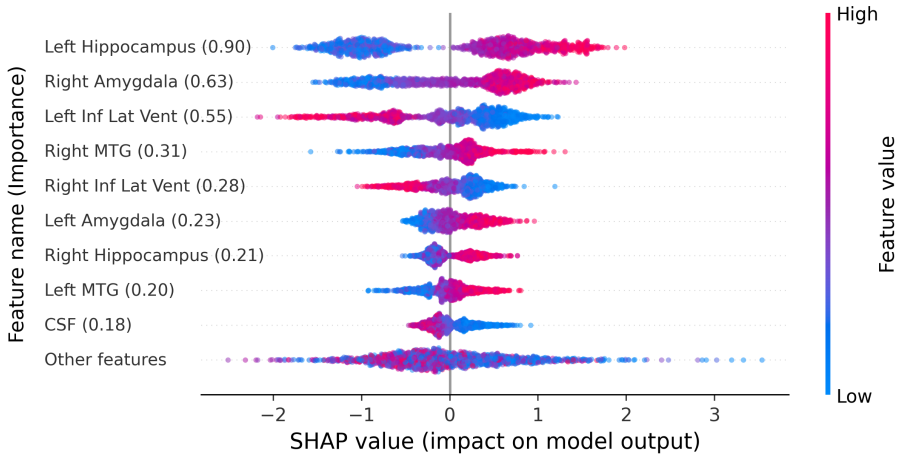

Fig. 3. SHAP analysis of the proposed regression model: red values on the positive axis indicate that higher volumes in specific brain regions are associated with healthier outcomes, whereas blue values denote the opposite relationship. (Abbreviations: Inf Lat Vent: inferior lateral ventricle; MTG: middle temporal gyrus; CSF: cerebrospinal fluid.)

while enhancing early-stage prediction capabilities through innovative loss function design. Our approach not only achieves state-of-the-art performance but also provides an extensible framework for integrating with existing survival analysis methods, thereby improving their long-term predictive accuracy. The experimental results demonstrate significant improvements in both overall accuracy and the reliability of early-stage predictions, which are crucial for timely clinical interventions. Further interpretation of the model reveals the most important features captured by the model, providing insights into neurodegenerative disease research. This advancement has important implications for Alzheimer's disease management and broader applications in medical research involving longitudinal survival analysis.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China (grant numbers 82441023, U23A20295, 62131015, 82394432, 6223355), the China Ministry of Science and Technology (S20240085, STI2030-Major Projects-2022ZD0209000, STI2030-Major Projects-2022ZD0213100), Shanghai Municipal Central Guided Local Science and Technology Development Fund (No. YDZX20233100001001, 24YF2729100), The Key R&D Program of Guangdong Province, China (grant number 2023B0303040001), and HPC Platform of ShanghaiTech University.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. De Strooper, B., Karran, E.: The cellular phase of Alzheimer's disease. *Cell* **164**(4), 603–615 (2016)
2. Liu, M., Zhang, D., Adeli, E., Shen, D.: Inherent structure-based multiview learning with multitemplate feature representation for alzheimer's disease diagnosis. IEEE Transactions on Biomedical Engineering **63**(7), 1473–1482 (2015)
3. Jang, J., Hwang, D.: M3t: three-dimensional medical image classifier using multi-plane and multi-slice transformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20718–20729 (2022)
4. Runde, B.S., Alapati, A., Bazan, N.G.: The optimization of a natural language processing approach for the automatic detection of alzheimer's disease using gpt embeddings. Brain Sciences **14**(3), 211 (2024)
5. Chintapalli, S.S., Govindarajan, S.T., Shou, H., Huang, H., Davatzikos, C.: Parsing disease heterogeneity using normative modelling and generative adversarial networks (gans). Alzheimer's & Dementia **20**, e094225 (2024)
6. Doody, R.S., Pavlik, V., Massman, P., Rountree, S., Darby, E., Chan, W.: Predicting progression of alzheimer's disease. Alzheimer's research & therapy **2**, 1–9 (2010)
7. Chong, M.S., Sahadevan, S.: Preclinical alzheimer's disease: diagnosis and prediction of progression. The Lancet Neurology **4**(9), 576–579 (2005)
8. Abbasi, A.F., Asim, M.N., Ahmed, S., Vollmer, S., Dengel, A.: Survival prediction landscape: an in-depth systematic literature review on activities, methods, tools, diseases, and databases. Frontiers in Artificial Intelligence **7**, 1428501 (2024)
9. Wiegrebe, S., Kopper, P., Sonabend, R., Bischl, B., Bender, A.: Deep learning for survival analysis: a review. Artificial Intelligence Review **57**(3), 65 (2024)
10. Lee, C., Zame, W., Yoon, J., Van Der Schaar, M.: Deephit: A deep learning approach to survival analysis with competing risks. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
11. Kopper, P., Wiegrebe, S., Bischl, B., Bender, A., Rügamer, D.: Deeppamm: Deep piecewise exponential additive mixed models for complex hazard structures in survival analysis. In: Pacific-Asia conference on knowledge discovery and data mining. pp. 249–261. Springer (2022)
12. Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Y.: Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. BMC medical research methodology **18**, 1–12 (2018)
13. Shahin, A.H., Zhao, A., Whitehead, A.C., Alexander, D.C., Jacob, J., Barber, D.: Centime: Event-conditional modelling of censoring in survival analysis. Medical Image Analysis **91**, 103016 (2024)
14. Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., et al.: The alzheimer's disease neuroimaging initiative (adni): Mri methods. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine **27**(4), 685–691 (2008)
15. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4itk: improved n3 bias correction. IEEE transactions on medical imaging **29**(6), 1310–1320 (2010)
16. Liu, F., Huang, J., Guo, L., Tang, H., Cai, X., Zhang, Y., Liu, J., Hua, R., Gu, J., Tao, T., et al.: Harmonizing multi-modality biases in infant development analysis with an integrated mri data processing pipeline. In: International Society for Magnetic Resonance in Medicine (2024)

17. Doshi, J., Erus, G., Ou, Y., Resnick, S.M., Gur, R.C., Gur, R.E., Satterthwaite, T.D., Furth, S., Davatzikos, C., Initiative, A.N., et al.: Muse: Multi-atlas region segmentation utilizing ensembles of registration algorithms and parameters, and locally optimal atlas selection. Neuroimage **127**, 186–195 (2016)
18. Johnson, W.E., Li, C., Rabinovic, A.: Adjusting batch effects in microarray expression data using empirical bayes methods. Biostatistics **8**(1), 118–127 (2007)
19. Authors, T.S.: Skopt: Scikit-optimize. Retrieved from https://scikit-optimize.github. io/ (2024)
20. Harrell Jr, F.E., Lee, K.L., Mark, D.B.: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine **15**(4), 361–387 (1996)
21. Lambert, J., Chevret, S.: Summary measure of discrimination in survival models based on cumulative/dynamic time-dependent roc curves. Statistical methods in medical research **25**(5), 2088–2102 (2016)
22. Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I.: From local explanations to global understanding with explainable ai for trees. Nature Machine Intelligence **2**(1), 2522–5839 (2020)
23. Aël Chetelat, G., Baron, J.C.: Early diagnosis of alzheimer's disease: contribution of structural neuroimaging. NeuroImage **18**(2), 525–541 (2003)
24. Veitch, D.P., Weiner, M.W., Aisen, P.S., Beckett, L.A., DeCarli, C., Green, R.C., Harvey, D., Jack Jr, C.R., Jagust, W., Landau, S.M., et al.: Using the alzheimer's disease neuroimaging initiative to improve early detection, diagnosis, and treatment of alzheimer's disease. Alzheimer's & Dementia **18**(4), 824–857 (2022)