Received: 24 October 2021 | Accepted: 1 August 2022

DOI: 10.1111/biom.13734

BIOMETRIC PRACTICE

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Neural network on interval-censored data with application to the prediction of Alzheimer's disease

Tao Sun[1,2] | **Ying Ding**[2]

[1]Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, China

[2]Department of Biostatistics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Correspondence
Tao Sun, Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, 100872, China.
Email: sun.tao@ruc.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 72101261, 72141306; National Bureau of Statistics of China, Grant/Award Number: 2021LZ18; Fund for building world-class universities (disciplines) of Renmin University of China, Grant/Award Number: KYGJC2021014; Ministry of Education of China, Grant/Award Number: 20JZD023; Public Health & Disease Control and Prevention, Major Innovation & Planning Interdisciplinary Platform for the "Double-First Class" Initiative of Renmin University of China

Abstract

Alzheimer's disease (AD) is a progressive and polygenic disorder that affects millions of individuals each year. Given that there have been few effective treatments yet for AD, it is highly desirable to develop an accurate model to predict the full disease progression profile based on an individual's genetic characteristics for early prevention and clinical management. This work uses data composed of all four phases of the Alzheimer's Disease Neuroimaging Initiative (ADNI) study, including 1740 individuals with 8 million genetic variants. We tackle several challenges in this data, characterized by large-scale genetic data, interval-censored outcome due to intermittent assessments, and left truncation in one study phase (ADNIGO). Specifically, we first develop a semiparametric transformation model on interval-censored and left-truncated data and estimate parameters through a sieve approach. Then we propose a computationally efficient generalized score test to identify variants associated with AD progression. Next, we implement a novel neural network on interval-censored data (NN-IC) to construct a prediction model using top variants identified from the genome-wide test. Comprehensive simulation studies show that the NN-IC outperforms several existing methods in terms of prediction accuracy. Finally, we apply the NN-IC to the full ADNI data and successfully identify subgroups with differential progression risk profiles. Data used in the preparation of this article were obtained from the ADNI database.

KEYWORDS
Alzheimer's disease prediction, genome-wide association studies, interval censoring, left truncation, neural network

1 | INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative and progressive disorder that accounts for about 70% of cases of dementia. It affects about 44 million individuals globally and causes about 4.6 million new cases each year. AD is known as a polygenic disorder, and the heritability of AD is estimated to be up to 79% (Gatz et al., 2006). Therefore, it has been of great interest in building prediction models for AD using genetic data. Moreover, because there are few effective treatments for AD, it is highly desirable to make early predictions for AD, enabling clinicians and patients to improve their quality of life before and during AD progression. The development of such a prediction model usually involves two steps: (1) identify genetic risk variants associated with AD through genome-wide association studies (GWAS) and (2) build a prediction model based on top variants from GWAS together with clinically

important predictors. To date, most GWAS studies on AD are looking for genetic variants associated with the onset of AD based on case-control studies (Jansen et al., 2019). However, it is still not well investigated the genetic causes underpinning AD development (i.e., from non-AD to AD), which is critical for early prevention and treatment. Moreover, to accommodate the high-dimensional genetic variants, most existing prediction models use a scalar polygenic risk score (PRS), a weighted linear summation of top variants. Despite its simplicity, PRS ignores the complex and nonlinear relationships among the high-dimensional genetic variants. In this paper, our goals are to (1) identify single-nucleotide polymorphisms (SNPs) associated with AD development (i.e., time-to-AD) through GWAS and (2) build an early and accurate prediction model for AD that can effectively extract nonlinear effects of the high-dimensional genetic risk variants.

This study is motivated by a complete database from all the four phases of the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Mueller et al., 2005), including 1740 Caucasian subjects with about 8 million SNPs. ADNI is a longitudinal study designed for the detection and development of AD. ADNI has gone through four phases: ADNI1 began in 2004, and the enrolled subjects are continuously followed together with new participants in three consecutive phases (ADNIGO, ADNI2, ADNI3). One major statistical challenge in analyzing the ADNI data is that the time-to-AD is not precisely observed due to intermittent assessment times, leading to interval-censored time-to-AD data. Further complications include right censoring, in which some subjects are still free of AD at the last assessment time, and left censoring, in which some subjects are enrolled with AD. Therefore, the occurrence time of AD is under general interval censoring, including a mixture of left-, right-, and interval censoring. Another statistical challenge is left truncation since subjects with AD are not included in the phase of ADNIGO. Ignoring interval censoring or left truncation may lead to biased model estimation and invalid inference results.

For regression models on interval-censored and left-truncated data, most existing works consider the proportional hazards assumption (e.g., Alioum & Commenges, 1996; Pan & Chappell, 2002; Shen, 2014; Gao & Chan, 2019; Wang et al., 2021). Recently, Shen et al. (2019) develop a semiparametric transformation model that includes a broad class of regression models and estimates parameters using a nonparametric maximum likelihood estimation approach. For survival/progression prediction models using interval-censored data with high-dimensional covariates (such as thousands of top SNPs from GWAS), several endeavors have been undertaken. For example, Li et al. (2020) develop a Cox model with adaptive Lasso under interval censoring. Wu et al. (2020) propose a penalized Cox model under interval censoring and utilize Bernstein polynomials to approximate nonlinear covariate effects. Yao et al. (2021) develop a survival forest method for interval-censored data. Recently, Sun et al. (2020) develop a neural network method for survival prediction under right censoring, achieving significantly better accuracy than several other methods (i.e., LASSO, random forest). In particular, its partial-likelihood-based loss function contains an unknown covariate-dependent function, approximated by a neural network inputted with high-dimensional covariates. One advantage of the neural network is that it can handle high-dimensional features and extract useful information from complex feature structures. Therefore, it is worth investigating whether the synergy of the neural network and genetic data can enhance accuracy in the progression prediction of AD, which is subject to interval censoring in ADNI. However, to be best of our knowledge, there is no neural network method for interval-censored data. One major difficulty lies in the simultaneous estimation of two unknown and nonlinear functions: the covariate-dependent function and the infinite-dimensional baseline hazard function. Moreover, as the baseline hazard is a function of time instead of covariates, the conventional neural network inputted with covariates cannot be applied here. Therefore, developing and implementing neural networks under interval censoring is more complicated than under right censoring.

In this paper, we first build a semiparametric transformation model for interval-censored and left-truncated data. Specifically, we develop a sieve estimation approach based on Bernstein polynomials and a computationally efficient score test for screening large amounts of covariates in settings like GWAS. Following that, we propose a novel neural network method for interval-censored data (NN-IC) that can simultaneously estimate the covariate-dependent function and the infinite-dimensional baseline hazard function. Specifically, we construct a new neural network based on Bernstein polynomials ("BPNet") to estimate the baseline hazard function. Finally, we apply these proposed methods to perform GWAS and develop the progression prediction model for AD using the ADNI data.

The remainder of this article is organized as follows. Section 2 describes the ADNI study, the current state-of-the-art prediction models for AD using ADNI, and how we process the data. In Section 3, we present our semiparametric transformation model for the interval-censored and left-truncated data. In Section 4, we introduce the NN-IC method, including its loss function, network architecture, hyperparameters, prediction evaluation, and interpretation. In Section 5, we examine our methods under various

simulation settings. In Section 6, we present the GWAS and progression prediction results in ADNI. Section 7 summarizes findings and discussions. The Supporting Information includes Web tables and figures.

2 | ADNI DATA AND EXISTING WORKS

Our data are obtained from the ADNI (Mueller et al., 2005). ADNI is a longitudinal multicenter study designed to develop clinical, imaging, and genetic biomarkers for the detection and development of AD. ADNI has recruited over 2300 individuals, consisting of people with cognitive normal (CN), mild cognitive impairment (MCI), and AD. The enrolled subjects were intermittently assessed during follow-up. The data used in this study were downloaded on May 1, 2021.

ADNI is one of the most well-known studies for predicting AD development. For example, Li et al. (2019) constructs a Cox-LASSO model with 256 MRI features extracted by neural networks from 2000 MCI subjects, obtaining better accuracy than the Cox model using traditional shape and texture features in MRI. Kong et al. (2018) develop a functional linear Cox regression model with scalar and time-varying predictors and use it to predict AD development in 373 MCI subjects. Li and Luo (2019) extract features from multiple time-varying predictors using multivariate functional principal component analysis (MFPCA) and apply them to a Cox model with 384 MCI individuals. Jiang et al. (2021) also extract features using MFPCA and applies them to an ensemble survival tree with 302 MCI subjects. Lin et al. (2021) extract features from multiple time-varying predictors and apply them to a random survival forest with 511 MCI individuals, outperforming the Cox model. Nakagawa et al. (2020) develop a deep learning survival model with gray matter volumes of 246 brain regions in 2000 CN/MCI subjects, outperforming the Cox model. However, all these studies convert interval-censored time-to-AD observations to right-censored data (i.e., using the observed interval's midpoint as the event time), resulting in potentially suboptimal prediction results. Additionally, these prediction models rarely use ADNI's GWAS data.

Moreover, most existing works using ADNI only utilize information from MCI individuals (45.4% of all participants), ignoring the individuals who were either CN or AD when entering ADNI. MCI is an intermediate disease stage between CN and AD. It has been well recognized that the underlying pathology of AD occurs before MCI (Petersen, 2009). Therefore, instead of just focusing on MCI patients, we make early predictions of AD in the general CN/MCI population to improve their quality of life before and during disease progression. Because ADNI recruits subjects at age 55 or above and AD rarely occurs before age 55 (Reitz et al., 2020), we set age 55 as the start time and calculate our event of interest (i.e., time-to-AD) on the age scale. One benefit of using age 55 as the start time is to make early and full-time course predictions in the population starting from age 55, when most individuals are free of AD pathology. Moreover, in aging-related studies, because aging is associated with a higher risk of chronic diseases like AD, age represents a more natural timescale than the study entry scale for characterizing the risk of AD (Lee et al., 2017). The third benefit of using the age scale instead of the study-entry scale is that it enables us to incorporate individuals who entered ADNI with AD as left-censored observations. Therefore, our work utilizes information from ADNI individuals with CN, MCI, or AD.

We use 1740 Caucasian individuals from all four phases of ADNI with genetic data and complete information about age, gender, education, and the *APOE* allele variant. The entire cohort can be divided into three groups: (a) those who developed AD before enrollment; (b) those who entered the study without AD and never developed AD during the follow-up; (c) the rest who entered the study without AD and later developed AD during the follow-up. For subjects in groups (a) and (b), the time-to-AD is left- and right-censored, respectively. For group (c), the time-to-AD is interval-censored (i.e., the AD event time lies between two adjacent assessment times). In addition, a left-truncation issue arises in the phase of ADNIGO because subjects with AD (i.e., group (a)) were not recruited.

Because the raw genetic data are obtained from different genotyping platforms across phases, we impute them to a common reference panel (1000 Genomes Project Phase 3 Release 5) through the Michigan Imputation Server. For quality control, we keep SNPs with Minimac3 $R^2 \geq 0.3$, MAF ≥ 0.01, Hardy–Weinberg equilibrium p-values ≥ 0.001, and missing genotype rate ≤ 0.05, leading to a total of 7,726,012 SNPs for each individual.

We randomly split the complete ADNI data into Data1 ($n = 1305$) and Data2 ($n = 435$) by a ratio of 3:1. The splitting is stratified based on the censoring status and study phases. We will use Data1 to perform GWAS, train and validate prediction models internally and use Data2 to validate prediction models externally. The characteristics of the complete data, Data1 and Data2 are summarized in Table 1.

TABLE 1 Characteristics of study subjects in the complete ADNI data and the randomly split Data1 and Data2

	Complete ($n = 1740$)	Data1 ($n = 1305$)	Data2 ($n = 435$)
Age			
Mean (SD)	73.5 (7.1)	73.4 (7.2)	74.0 (6.9)
Median (range)	73.5 (55.0-91.4)	73.5 (55.0-91.4)	73.5 (55.0-90.3)
Gender (n, %)			
Female	780 (44.8%)	587 (45.0%)	193 (44.4%)
Male	960 (55.2%)	718 (55.0%)	242 (55.6%)
Education			
Mean (SD)	16.1 (2.8)	16.1 (2.8)	15.9 (2.7)
Median (range)	16 (4-20)	16 (4-20)	16 (7-20)
APOE (n, %)			
Zero allele	941 (54.1%)	697 (53.4%)	244 (56.1%)
One allele	637 (36.6%)	487 (37.3%)	150 (34.5%)
Two alleles	162 (9.3%)	121 (9.3%)	41 (9.4%)
Censoring types (n, %)			
Left-censored	300 (17.2%)	222 (17.0%)	78 (17.9%)
Interval-censored	342 (19.7%)	257 (19.7%)	85 (19.6%)
Right-censored	1098 (63.1%)	826 (63.3%)	272 (62.5%)
Study (n, %)			
ADNI1	703 (40.4%)	537 (41.2%)	166 (38.2%)
ADNIGO	114 (6.5%)	84 (6.4%)	30 (6.9%)
ADNI2	619 (35.6%)	462 (35.4%)	157 (36.1%)
ADNI3	304 (17.5%)	222 (17.0%)	82 (18.8%)
Follow-up years			
Mean (SD)	3.1 (3.2)	3.1 (3.2)	3.0 (3.3)
Median (range)	2.0 (0.0-15.0)	2.0 (0.0-14.7)	2.0 (0.0-15.0)

3 | SEMIPARAMETRIC TRANSFORMATION MODEL FOR THE INTERVAL-CENSORED AND LEFT-TRUNCATED DATA

3.1 | Notations and assumptions

First, we define notations for the interval-censored and left-truncated data as in ADNIGO. Denote T as the time from the initial event (i.e., age 55) to the event of interest (i.e., the occurrence of AD), A as the truncation time from the initial event to the study entry, and $\boldsymbol{Z}$ as the covariate vector. In the presence of left truncation, only subjects who satisfy $A \le T$ are observed. Let $(\widetilde{A}, \widetilde{T}, \widetilde{\boldsymbol{Z}})$ denote the realization of $(A, T, \boldsymbol{Z})$ given $A \le T$. Let $\boldsymbol{Q} = (Q_1, \dots, Q_K)^T$ be K gap times between two assessments following the study entry and define $(\widetilde{U}_1, \dots, \widetilde{U}_K)$ as the assessment times after the enrollment, where $\widetilde{U}_k = \widetilde{A} + \sum_{l=1}^{k} Q_k, k = 1, \dots, K$. Since ADNI individuals follow a prespecified assessment time schedule, it is reasonable to assume that $\boldsymbol{Q}$ is independent with $(\widetilde{A}, \widetilde{T})$ given $\widetilde{Z}$. When $\widetilde{T} \in (\widetilde{U}_l, \widetilde{U}_{l+1}]$ for $l \in \{1, \dots, K-1\}$, we write the observed time interval for $\widetilde{T}$ as $(\widetilde{L}, \widetilde{R}] = (\widetilde{U}_l, \widetilde{U}_{l+1}]$. When $\widetilde{T} < \widetilde{U}_1$, we have $(\widetilde{L}, \widetilde{R}] = (\widetilde{A}, \widetilde{U}_1]$. When $\widetilde{T} > \widetilde{U}_K$, we have $(\widetilde{L}, \widetilde{R}] = (\widetilde{U}_K, \infty]$ corresponding to right censoring. Therefore, the observed data for the ith subject in ADNIGO are written as $\boldsymbol{D}_{1i} = \{\widetilde{A}_i, \widetilde{L}_i, \widetilde{R}_i, \widetilde{\boldsymbol{Z}}_i\}$, where $i = 1, \dots, n_1$ for n_1 independently and identically distributed (i.i.d.) subjects.

For the rest phases without the left-truncation issue, the observation for the ith subject is denoted as $\boldsymbol{D}_{2i} = \{L_i, R_i, \boldsymbol{Z}_i\}$, where $i = 1, \dots, n_2$ for n_2 i.i.d. subjects, L_i and R_i form the observed time interval for the event time T_i and $\boldsymbol{Z}_i$ is a p-dimension covariate vector.

For the regression model of T given covariates $\boldsymbol{Z}$, we consider the semiparametric transformation model, with the survival function expressed as

$$S(t|\boldsymbol{Z}) = \exp[-G\{\exp(\boldsymbol{Z}^T\boldsymbol{\beta})\Lambda(t)\}], \tag{1}$$

where $G(\cdot)$ is a prespecified increasing function, $\boldsymbol{\beta}$ is the regression coefficient vector, and $\Lambda(t)$ is a nondecreasing function of time t. We choose the logarithmic function $G(x) = \log(1 + rx)/r$ that includes proportional hazards (PH; $r = 0$) and proportional odds (PO; $r = 1$) models.

Therefore, the transformation model is more flexible than the traditional PH or PO model.

3.2 | Joint likelihood for ADNI individuals

We build the joint likelihood for individuals from all four phases of ADNI. We assume all individuals share the same parameters $(\boldsymbol{\beta}, \Lambda)$. For ADNIGO with the left truncation issue, we first examine whether the truncation time satisfies the length-biased assumption, which is rejected with a p-value of 0.02 by a formal test (Addona & Wolfson, 2006). Therefore, we adopt the conditional likelihood approach commonly used for left-truncated data, assuming that A and T are independent given the covariates $\boldsymbol{Z}$. The conditional likelihood of $(\widetilde{L}, \widetilde{R})$ given $(\widetilde{A}, \widetilde{\boldsymbol{Z}})$ for ADNIGO individuals can be written as

$$L^C_{n_1}(\boldsymbol{\beta}, \Lambda|\boldsymbol{D}_1) = \prod_{i=1}^{n_1} \frac{\exp[-G\{\exp(\widetilde{\boldsymbol{Z}}_i^T\boldsymbol{\beta})\Lambda(\widetilde{L}_i)\}] - \exp[-G\{\exp(\widetilde{\boldsymbol{Z}}_i^T\boldsymbol{\beta})\Lambda(\widetilde{R}_i)\}]}{\exp[-G\{\exp(\widetilde{\boldsymbol{Z}}_i^T\boldsymbol{\beta})\Lambda(\widetilde{A}_i)\}]}. \tag{2}$$

For individuals from the rest phases, the full likelihood of (L, R) given $\boldsymbol{Z}$ can be written as

$$L^F_{n_2}(\boldsymbol{\beta}, \Lambda|\boldsymbol{D}_2) = \prod_{i=1}^{n_2} (\exp[-G\{\exp(\boldsymbol{Z}_i^T\boldsymbol{\beta})\Lambda(L_i)\}] - \exp[-G\{\exp(\boldsymbol{Z}_i^T\boldsymbol{\beta})\Lambda(R_i)\}]). \tag{3}$$

Overall, the joint likelihood for all ADNI individuals is expressed as

$$L_n(\boldsymbol{\beta}, \Lambda|\boldsymbol{D}_1, \boldsymbol{D}_2) = L^C_{n_1}(\boldsymbol{\beta}, \Lambda|\boldsymbol{D}_1) \times L^F_{n_2}(\boldsymbol{\beta}, \Lambda|\boldsymbol{D}_2). \tag{4}$$

3.3 | Sieve estimation and the generalized score test

In our joint likelihood (4), we are interested in estimating the unknown parameter $\theta \in \Theta$, where $\Theta = \{\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \Lambda)^T \in \mathcal{B} \otimes \mathcal{M}\}$, $\mathcal{B} = \{(\boldsymbol{\beta} \in R^p, \|\boldsymbol{\beta}\| \le M\}$ with M being a positive constant, and $\mathcal{M}$ the collection of all bounded, continuous, and nondecreasing functions over $[c, u]$. We need to estimate finite-dimensional parameters $\boldsymbol{\beta}$ and an infinite-dimensional parameter $\Lambda(t)$ simultaneously. Following Zhou et al. (2017), we use Bernstein polynomials to build a sieve space $\Theta_n = \{\boldsymbol{\theta}_n = (\boldsymbol{\beta}^T, \Lambda_n)^T \in \mathcal{B} \otimes \mathcal{M}_n\}$. Here, $\mathcal{M}_n$ is the space defined as

$$\mathcal{M}_n = \left\{ \Lambda_n(t) = \sum_{k=0}^{m_n} \phi_k B_k(t, m_n, c, u) : \sum_{k=0}^{m_n} |\phi_k| \le M_n; \right.$$
$$\left. 0 \le \phi_0 \le \cdots \le \phi_{m_n} \right\}, \tag{5}$$

where $B_k(t, m_n, c, u)$ represents the Bernstein basis polynomial defined as $B_k(t, m_n, c, u) = \binom{m_n}{k}\{(t-c)/(u-c)\}^k\{1-(t-c)/(u-c)\}^{m_n-k}$, with degree $m_n = o(n^\nu)$ for some $\nu \in (0, 1)$ and $k = 0, \ldots, m_n$. By maximizing $l_n(\boldsymbol{\theta}; D_1, D_2) = \log L_n(\boldsymbol{\theta}; D_1, D_2)$ over the sieve space Θ_n, we obtain the sieve estimators $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\beta}}_n^T, \widehat{\Lambda}_n)^T$. For the variance–covariance of $\widehat{\boldsymbol{\beta}}_n$, we invert the observed information matrix and take the corresponding block.

We aim to test millions of SNPs one by one. We propose a computationally efficient generalized score test. Specifically, we denote $\boldsymbol{\beta} = (\beta_g, \boldsymbol{\beta_{ng}})$, where β_g is the parameter of interest for testing (i.e., SNP) and $\boldsymbol{\beta_{ng}}$ are the rest coefficients. Then the null hypothesis for one single SNP is expressed as $H_0 : \beta_g = 0$ and $(\boldsymbol{\beta_{ng}}, \Lambda)$ is arbitrary. Denote $\widehat{\boldsymbol{\theta}}_0 = (\beta_g = 0, \widehat{\boldsymbol{\beta}}_{ng0}, \widehat{\Lambda}_0)$ as the restricted sieve maximum likelihood estimator under the null hypothesis. By following Cox and Hinkley (1979), we obtain the generalized score test statistics as

$$T_s = \boldsymbol{U}^T(\widehat{\boldsymbol{\theta}}_0)\boldsymbol{J}^{-1}(\widehat{\boldsymbol{\theta}}_0)U(\widehat{\boldsymbol{\theta}}_0), \tag{6}$$

where $\boldsymbol{U}(\widehat{\boldsymbol{\theta}}_0)$ and $\boldsymbol{J}(\widehat{\boldsymbol{\theta}}_0)$ are the score function and observed information matrix, respectively. The test statistics follow a χ^2 distribution with a degree of freedom being 1. One big advantage of the score test is that one only needs to estimate the model parameters once under the null model without any SNP, which greatly reduces computation time in GWAS (Sun et al., 2019).

4 | NEURAL NETWORK FOR INTERVAL-CENSORED DATA

We propose a novel neural network method for interval-censored data (NN-IC). We do not consider left truncation because we predict AD starting at age 55, at which AD rarely occurs. Moreover, we aim to predict AD risk for a new subject who does not necessarily enter ADNI, and thus the left-truncation issue does not apply.

4.1 | Assumption and loss function

Our NN-IC model is based on the PH assumption, the most popular assumption for censored data. Similar to Section 3, we assume the assessment times (i.e., L and R) are independent of the event time given covariates. The NN-IC model is expressed as $\Lambda(t|\boldsymbol{Z}_i) = \Lambda(t)e^{g(\boldsymbol{Z}_i;\theta)}$, where $\Lambda(t)$ is the unspecified baseline cumulative hazard function at time t, $g(\boldsymbol{Z_i}; \boldsymbol{\theta})$ is the prognostic index with an unknown form for the function $g(\cdot)$, and $\boldsymbol{\theta}$ is the parameter set. One major advantage of NN-IC compared to the regular

PH model is that NN-IC can estimate various nonlinear structures of $g(\cdot)$ using neural networks (Hornik et al., 1989), whereas the regular PH model simply assumes $g(\cdot)$ is a linear function, which is hardly true in the presence of high-dimensional covariates. To mitigate the overfitting issue in neural networks, we follow existing works (Bello et al., 2019) to apply the L_1 penalty to the loss function $-l(\theta, \Lambda; \boldsymbol{Z}) + \lambda\|\theta\|_1$, where $l(\theta, \Lambda; \boldsymbol{Z})$ is the log-likelihood function:

$$\frac{1}{n}\sum_{i=1}^{n} \log[\exp\{-\Lambda(L_i)e^{g(\boldsymbol{Z}_i;\theta)}\} - \exp\{-\Lambda(R_i)e^{g(\boldsymbol{Z}_i;\theta)}\}]. \quad (7)$$

4.2 | NN-IC architecture and hyperparameters

The loss function of NN-IC is essentially a semiparametric function, involving unspecified infinite-dimensional parameters $(\Lambda(L_i), \Lambda(R_i))$ as well as the unknown covariate-dependent function $g(\boldsymbol{Z}_i; \theta)$. We need to estimate all these functions using neural networks simultaneously. However, the conventional method that approximates a function with a neural network inputted with covariates does not work for estimating $\Lambda(L_i)$ and $\Lambda(R_i)$. Therefore, the estimation of NN-IC is more complicated than the usual existing partial-likelihood-based neural network methods (Faraggi & Simon, 1995; Katzman et al., 2018).

To solve the estimation problem in NN-IC, we propose a novel multiple neural-network-based estimation approach. First, for the estimation of $g(\boldsymbol{Z}_i; \theta)$, we use a regular L-hidden-layer neural network inputted with covariates. Specifically, $\boldsymbol{Z}_i$ constitutes the input nodes of the neural network, the parameter θ represents the collection of all weights in the network, and the output of the network is $\widehat{g}(\boldsymbol{Z}_i; \widehat{\theta})$. Second and more importantly, for the estimation of $\Lambda(\cdot)$, we propose a new type of neural network constructed by Bernstein polynomials, named as "BPNet" and build two separate BPNets to approximate $\Lambda(L_i)$ and $\Lambda(R_i)$, respectively. Take the BPNet for $\Lambda(L_i)$ as an example. The BPNet takes L_i as its input node, connects the input node with a hidden layer containing $(m_n + 1)$ number of hidden nodes and produces a scalar value o_i in the output node. Specifically, for the kth hidden node a_k $(k = 0, \ldots, m_n)$, we have $a_k = f_k(L_i)$, where f_k is the activation function for a_k, and it takes the form of a Bernstein basis polynomial as defined in Section 3, that is $a_k = f_k(L_i) = B_k(L_i, m_n, c, u)$. For the output node o_i of this BPNet, it can be expressed as

$$o_i = f^{out}\Big(\sum_{k=0}^{m_n} w_k a_k\Big) = \sum_{k=0}^{m_n} w_k a_k = \sum_{k=0}^{m_n} w_k B_k(L_i, m_n, c, u), \quad (8)$$

where f^{out} is an identity function, w_k is the weight parameter satisfying $0 \le w_0 \le \cdots \le w_{m_n}$. Typically we have $o_i = \Lambda_n(L_i)$. The parameter set in this BPNet is composed of $\{w_k, k = 0, \ldots, m_n\}$. Moreover, the BPNets for $\Lambda_n(L_i)$ and $\Lambda_n(R_i)$ share the same set of parameters w_k. Overall, the full parameter set in NN-IC is composed of $\{\theta, w_k, k = 0, \ldots, m_n\}$. By maximizing the loss function in Equation 7, we can obtain the estimators $\widehat{\theta}_n$ and $\widehat{\Lambda}_n$ (expressed by $\widehat{w}_k$). We use the mini-batch stochastic gradient descent algorithm for optimizing the loss function. Once we get $\widehat{g}(\boldsymbol{Z}_i; \widehat{\theta}_n)$ and $\widehat{\Lambda}_n$, we can obtain the predicted survival probability for subject i at time t through $\widehat{S}(t|\boldsymbol{Z}_i) = \exp\{-\widehat{\Lambda}_n(t)e^{\widehat{g}(\boldsymbol{Z}_i;\widehat{\theta}_n)}\}$.

The NN-IC method involves hyperparameter selection. For the neural network in $g(\boldsymbol{Z}_i; \theta)$ estimation, hyperparameters include the number of hidden layers, number of nodes per hidden layer, choice of activation function, the L_1 penalty parameter, batch size, epoch size, and learning rate. For BPNets, the hyperparameter is the degree of Bernstein polynomials. We select the hyperparameters using cross-validations as described in Section 4.3. For real data analysis, we use the following hyperparameters: one hidden layer, 50 nodes per hidden layer, activation function Scaled Exponential Linear Unit (SeLU), L_1 penalty $= 0.5$, batch size $N_B = 50$, epoch size $N_E = 1000$, learning rates 0.002, uniformly distributed initial values, and Bernstein polynomial degree $m_n = 5$. For the simulations, we select the following hyperparameters: two hidden layers, 50 nodes per hidden layer, activation function SeLU, L_1 penalty $= 0.5$, batch size 50, epoch size 1000, learning rate 0.01, uniformly distributed initial values, and $m_n = 3$.

4.3 | Prediction evaluation, validation, and interpretation

We use various metrics to evaluate prediction accuracy under interval censoring. For real data analysis, we use the integrated Brier score (IBS) (Tsouprou, 2015), expressed as

$$IBS(\widehat{S}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{u}\int_0^u \{I(T_i > t|\boldsymbol{Z}_i) - \widehat{S}(t|\boldsymbol{Z}_i)\}^2 dt, \quad (9)$$

where u is the maximum finite value of all observed $\{L_i, R_i\}$. When $R_i < t$, we have $I(T_i > t|\boldsymbol{Z}_i) = 0$. When $L_i \ge t$, we have $I(T_i > t|\boldsymbol{Z}_i) = 1$. When $L_i < t \le R_i$, the exact value of $I(T_i > t|\boldsymbol{Z}_i)$ is unknown. We estimate $I(T_i > t|\boldsymbol{Z}_i)$ by $\widehat{I}(T_i > t|\boldsymbol{Z}_i) = \{\widehat{S}(t|\boldsymbol{Z}_i) - \widehat{S}(R_i|\boldsymbol{Z}_i)\}/\{\widehat{S}(L_i|\boldsymbol{Z}_i) - \widehat{S}(R_i|\boldsymbol{Z}_i)\}$. In the special case when $L_i < t \le R_i = \infty$, $\widehat{I}(T_i > t|\boldsymbol{Z}_i) = \widehat{S}(t|\boldsymbol{Z}_i)/\widehat{S}(L_i|\boldsymbol{Z}_i)$. We also consider two additional evaluation metrics: the proportion of the predicted median survival time lying outside $(L_i, R_i]$ (denoted as p_{out}), as

well as the absolute distance of the predicted median time below L_i or above R_i when the predicted median time falls outside $(L_i, R_i]$ (denoted as d_{out}). In simulation studies where the true survival function and event time T are known, we employ the mean squared prediction error, which is essentially the average integrated L_2 distance between the true and the estimated survival functions, expressed as

$$L(\widehat{S}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\int_0^{T_i}\{S(t|\boldsymbol{Z}_i) - \widehat{S}(t|\boldsymbol{Z}_i)\}^2 dt. \quad (10)$$

Smaller values of these metrics indicate better prediction performance.

Overfitting is a common issue in developing prediction models. One way to alleviate the issue is to first select the optimal set of hyperparameters in the training dataset via internal cross-validation, build a final model with the selected hyperparameters in the entire training dataset, and evaluate it in an external dataset.

It is crucial to interpret the prediction of NN-IC. We employ the local interpretable model-agnostic explanation (LIME) method (Ribeiro et al., 2016), which calculates a prediction importance level for each predictor in each subject. More details about how to implement LIME in a deep learning survival model can be found in Sun et al. (2020).

5 | SIMULATION STUDIES

5.1 | Simulation I: Parameter estimation in interval-censored and left-truncated data

We evaluate the estimation performance of the semiparametric transformation model presented in Section 3. The true event times are generated from the PH model with Weibull baseline hazards (scale $\lambda = 0.1$ and shape $k = 2$) and the PO model with Loglogistic baseline hazards (scale $\lambda = 1$ and shape $k = 2$). Two nongenetic covariates are generated from a normal distribution $N(6, 2^2)$ and a Bernoulli distribution ($p = 0.5$), with regression coefficients $\beta_{ng1} = \beta_{ng2} = 0.1$. A genetic covariate, coded as 0 or 1 or 2, is generated from a multinomial distribution with probabilities $\{(1-p)^2, 2p(1-p), p^2\}$, where $p = 40\%$ is the minor allele frequency. The coefficient of the genetic covariate is set as $\beta_g = 0$. The truncation times are generated from an exponential distribution. To obtain interval-censored data, we followed the procedure in Sun and Ding (2021), which fits the study design of ADNI. Explicitly, we assume each subject is assessed for K times with the length between two adjacent assessments following an exponential distribution. For each subject i, $\widetilde{L}_{ij}$ is defined as the last assessment time before T_{ij} and $\widetilde{R}_{ij}$ is the first assessment time after T_{ij}. When T_{ij} is larger than the last assessment time, T_{ij} is right-censored at the last assessment time. The sample size is $n = 500$. We use the Bernstein polynomial degree $m_n = 3$. For the time range $[c, u]$, we choose $c = 0$ and set u as the largest value of all $\{\widetilde{L}_{ij}, \widetilde{R}_{ij}\}$ plus a constant. We repeat the simulations 1000 times and report the results in Web Table 1. Our sieve estimators are all unbiased, and all empirical coverage probabilities are close to the nominal level.

We compare the computing time of our proposed generalized score test with Wald and likelihood ratio tests. For screening 5000 genetic variants, the three tests take 2.5, 13, and 12.5 min. Therefore, the score test is about five times faster than the other tests. When screening 7.7 million genetic variants as in ADNI, the computation times are approximately 3 days, 14 days, and 13 days using the score, Wald, and likelihood ratio tests, respectively. Thus, the score test greatly enhances the computational efficiency in real applications.

5.2 | Simulation II: Survival prediction in interval-censored data

We evaluate the prediction performance of NN-IC together with two prediction methods for interval-censored data: the adaptive lasso model under the PH assumption ("ALASSO") (Li et al., 2020) and the conditional survival forest model ("ICcforest") (Yao et al., 2021). For ALASSO, we use the R package {*ALassoSurvIC*}, which searches the optimal tuning parameter automatically based on Bayesian information criterion (BIC). For ICcforest, we use the R package {*ICcforest*}, which provides two approaches for finding *mtry* (i.e., the number of randomly selected predictors at each split). One approach is based on the out-of-bag error estimates, and the other is to set $mtry = \sqrt{p}$. We find that the latter approach results in better prediction performance in the simulated and real data. Therefore, we use $mtry = \sqrt{p}$. For NN-IC, since we have many hyperparameters and multiple simulation scenarios, we fix one set of hyperparameters for all scenarios. More details about the NN-IC hyperparameters are described in Section 4.

In genetics and genomics data, we observe that many predictors have (nonzero) weak effects due to correlations among predictors. Therefore, we generate data with weak effects and set the number of predictors as $p = 20, 50, 100, 500$. We consider the following scenarios:

$$\text{Scenario1}: h(t|Z_i) = h_0(t)\exp\left(\sum_{j=1}^{p}\beta_j Z_{ij}\right), \quad (11)$$

$$\text{Scenario2}: h(t|Z_i) = h_0(t)\exp\left(\sum_{j=1}^{p}\beta_j Z_{ij} + Z_{i1}^2 + Z_{i2}^2\right), \tag{12}$$

$$\text{Scenario3}: h(t|Z_i) = h_0(t)\exp\left(\sum_{j=1}^{p}\beta_j Z_{ij} + Z_{i3}Z_{i4}\right), \tag{13}$$

$$\text{Scenario4}: h(t|Z_i)$$
$$= h_0(t)\exp\left(\sum_{j=1}^{p}\beta_j Z_{ij} + Z_{i1}^2 + Z_{i2}^2 + Z_{i3}Z_{i4}\right), \tag{14}$$

$$\text{Scenario5}: h(t|Z_i)$$
$$= h_0(t)\exp\left(\sum_{j=1}^{p}\beta_j Z_{ij} + I(Z_{i1} < -0.5 \cup Z_{i2} < -0.5)\right.$$
$$\left. - I(Z_{i1} \geq -0.5 \cap Z_{i2} \geq -0.5) + Z_{i3}Z_{i4}\right), \tag{15}$$

where $h_0(t)$ is the baseline Weibull hazard function with $\lambda = 0.01, k = 10$. We generate Z_i from a multivariate normal distribution $MVN(0, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = \{\sigma_{jj'} = e^{-|j-j'|}, 1 \leq j, j' \leq p\}$. Then the first 20% Z_{ij} remain continuous, the second 20% Z_{ij} are transformed into binary predictors through $I(Z_{ij} > 0)$, and the rest 60% Z_{ij} are transformed into multinomial predictors through $I(Z_{ij} > -0.5) + I(Z_{ij} > 0.5)$. We set $\beta_j = 0.2$ for continuous and binary predictors. For multinomial predictors, we mimic the linkage disequilibrium effect in SNP data by generating β_j from $MVN(0.2, 0.01 \times \mathbf{\Sigma})$. The right-censoring rates are 50%, and sample sizes are 1000. We train the models in a training dataset, test them in a test dataset, and summarize the results across 200 replications. In Table 2, we compare NN-IC, ICcforest, and ALASSO in terms of the mean squared prediction error between the true and predicted survival probabilities. A smaller value of the prediction error indicates better prediction accuracy. In general, NN-IC outperforms the other models. As p increases, all methods' performance declines. We also evaluate the three methods in data generated from a PO model and find that NN-IC outperforms the other methods, especially when p is large (Web Table 2). Therefore, NN-IC seems to have some robustness against the misspecification of the PH assumption.

We further evaluate the effects of sample sizes, number of hidden layers, number of nodes per layer, and choices of initial parameter values on the prediction performance of NN-IC. We choose scenario 5 with $p = 500$ and present the results in Web Figure 1. Overall, the mean squared prediction error decreases as the sample size increases, and the increment is more obvious between smaller sample sizes, such as from $n = 200$ to 500 or $n = 500$ to 1000. This demonstrates that NN-IC requires a moderately large sample size to achieve good prediction performance when the number of predictors is relatively large. The performance of NN-IC is relatively stable using different hidden layers, suggesting that NN-IC with a few hidden layers shall be sufficient. For the number of nodes per layer, when the number reaches 75 or above, the mean squared prediction error steadily increases, suggesting a moderate number of nodes (i.e., 50) is appropriate. Lastly, the choices of initial values affect NN-IC's performance, with the lowest mean squared prediction error corresponding to initial values sampled from a uniform distribution between 0 and 1 (which is what we use in all analyses).

TABLE 2 The mean squared prediction error (×100) averages and standard deviations (SD) from 200 replications for the NN-IC, ICcforest, and ALASSO models under five scenarios: linear effects (scenario 1) and linear effects together with nonlinear effects (scenario 2) or with interactions (scenario 3) or with nonlinear and interaction effects (scenario 4) or with interaction and indicator effects (scenario 5). The number of predictors is set at $p = 20, 50, 100, 500$

	p	NN-IC	ICcforest	ALASSO
Scenario 1	20	0.7 (0.3)	0.7 (0.3)	0.5 (0.4)
	50	1.1 (0.6)	1.7 (0.9)	1.0 (0.8)
	100	1.9 (1.7)	2.9 (1.5)	2.1 (1.8)
	500	3.0 (4.0)	6.4 (3.4)	8.0 (3.5)
Scenario 2	20	0.8 (0.8)	1.0 (0.4)	1.0 (0.4)
	50	1.4 (1.2)	1.9 (1.0)	1.4 (0.9)
	100	1.8 (1.4)	3.0 (1.5)	2.8 (1.8)
	500	4.4 (7.1)	6.5 (3.4)	8.2 (3.4)
Scenario 3	20	0.8 (0.4)	0.9 (0.4)	0.8 (0.4)
	50	1.2 (0.6)	1.8 (0.9)	1.3 (0.9)
	100	1.8 (1.4)	3.0 (1.5)	2.4 (1.9)
	500	3.3 (2.8)	6.5 (3.4)	7.7 (3.4)
Scenario 4	20	0.8 (0.3)	1.2 (0.4)	1.2 (0.5)
	50	1.3 (0.7)	2.0 (1.0)	1.6 (0.9)
	100	2.0 (1.6)	3.1 (1.5)	2.7 (1.9)
	500	4.0 (3.7)	6.6 (3.4)	7.9 (3.3)
Scenario 5	20	0.7 (0.3)	0.9 (0.4)	0.8 (0.4)
	50	1.2 (0.6)	1.8 (0.9)	1.3 (0.8)
	100	1.8 (1.3)	2.9 (1.5)	2.5 (1.8)
	500	4.4 (4.7)	6.5 (3.4)	8.3 (3.4)

6 | APPLICATION TO ADNI DATA

6.1 | GWAS results

We analyze Data1 by applying the semiparametric transformation model in Section 3. We build two null models,

one using only clinical factors (gender, education, phase of ADNI), and the other adjusting for the genetic factor *APOE*. We choose the model with the smallest BIC, corresponding to $g(x) = \log(1 + 0.2x)/0.2$ and $m_n = 3$.

We perform GWAS in 7.7 million SNPs using the proposed generalized score test under both null models and plot the $-\log_{10}(p)$ values in Web Figure 2. For GWAS without adjusting *APOE*, multiple SNPs from the *PVRL2–TOMM40–APOE–APOC1* region on chromosome 19 reach the "genome-wide" significance level ($p < 5 \times 10^{-8}$). This region is significantly associated with AD onset in case-control studies (Jansen et al., 2019). In GWAS adjusting for *APOE*, we find one significant gene *CHRNA4* on chromosome 20, which is also known for AD (Kawamata & Shimohama, 2002). Moreover, we successfully identify several risk variants that have not been previously reported. For example, multiple SNPs in the *CDKN2AIP* gene on chromosome 4 reach the moderate significance level ($p < 1 \times 10^{-5}$) in both GWAS. The gene plays a central role in DNA damage response and influences multiple signaling pathways involved in cell proliferation, apoptosis, and senescence (Cheung et al., 2014). These new findings may contribute to the understanding of AD development.

Recent works suggest that SNPs with large p-values can still contribute to disease prediction (Escott-Price et al., 2017). Therefore, we relax the p-value threshold to $p < 1 \times 10^{-3}$. We perform SNP clumping (Privé et al., 2018) to extract representative SNPs out of the top SNPs, and obtain 71, 371, 623, and 1970 SNPs at the p-value thresholds of $1 \times 10^{-5}, 1 \times 10^{-4}, 2 \times 10^{-4}$, and 1×10^{-3} respectively. We use these representative SNPs and nongenetic predictors (i.e., gender, education) to develop AD prediction models in Section 6.2.

6.2 | Prediction model comparisons and results

We apply and compare several methods to make progression predictions for AD, including (i) the regular semiparametric PH model ("icenReg") using the R package *icenReg* (Anderson-Bergman, 2017), (ii) the Bayesian semiparametric PH model ("Bayesian") (Lin et al., 2015), (iii) the adaptive lasso model ("ALASSO") (Li et al., 2020), (iv) the conditional survival forest model ("ICcforest") (Yao et al., 2021), and (v) our proposed NN-IC method. We perform fivefold cross-validation, train the models in Data1, and evaluate their predictions in Data2. For ALASSO and ICcforest, we use their optimal parameter tuning procedure as described in Section 5.2. For NN-IC, we select the set of hyperparameters that gives the best average prediction performance (i.e., IBS) across the five replications. The final choice of NN-IC hyperparameters is described in Section 4. We also include two benchmark models ("APOE" and "PRS"), which are standard PH models using gender, education, and *APOE* or the PRS as predictors. Specifically, PRS is a weighted sum of the 31 AD-associated SNPs reported in Desikan et al. (2017). The weights are from a survival model analyzing time-to-AD onset based on a cohort not included in ADNI.

We compare the prediction performance (i.e., using metrics such as IBS, p_{out}, d_{out}) across different models in multiple scenarios corresponding to varying p-value thresholds from GWAS. Smaller values of these evaluation metrics represent better prediction performance. For the internal cross-validation in Data1 (Table 3, panel a), the performance of NN-IC improves as the number of predictors increases among the four scenarios presented. We also check an additional scenario with more SNPs (e.g., 3629 SNPs at the threshold of $p < 2 \times 10^{-3}$) and find that the performance of NN-IC decreases compared to the performance with 1970 SNPs in scenario 4 (results not shown). For the external validation in Data2 (Table 3, panel b), NN-IC generally achieves better prediction performance than the other methods, particularly in scenario 4. The computing times of NN-IC for fitting Data1 once are 16–24 s under scenarios 1–4. In addition, the icenReg, Bayesian, and ALASSO methods fail due to difficulties in handling large matrices when the number of SNPs is moderately large. Although the Bayesian method seems to give a better prediction than NN-IC in Data1 in scenario 1, it performs less optimally than NN-IC in Data2. The PRS model gives similar prediction metrics as the APOE model in the internal and external validations.

We also calculate the time-dependent Area under the ROC Curve (AUC) values (Wu et al., 2020) in Data2 for APOE, PRS, ICcforest, and NN-IC under scenario 4. As shown in Web Table 3, the AUC values from NN-IC are generally similar to or higher than the other models across different time points from age 60 to 80. Although NN-IC gives lower AUC than APOE and PRS at the late age of 80, its AUC values are higher in the early to middle ages. In addition, NN-IC shows advantages over APOE and PRS in terms of other evaluation metrics, as shown in Table 3.

To interpret predictions by NN-IC, we obtain the predictor importance measure for each subject in Data2 using the LIME method under scenario 4. One advantage of LIME is that it provides a subject-specific interpretation of predictor importance. Figure 1A illustrates the top 10 important predictors, among which some are harmful (red) or protective (blue). In particular, the minor allele of *rs*429358 in the *APOE* gene shows the strongest harmful effect for AD in all subjects. We also plot the top 10 important predictors without *APOE* in Figure 1B. We find that one genetic variant could be important for some individuals but not for others (visualized by different vertical color bands

TABLE 3 The fivefold internal cross-validation (a) and external validation (b) results from seven prediction models (APOE, PRS, icenReg, Bayesian, ALASSO, ICcforest, NN-IC). Scenarios 1-4 represents different p-value thresholds ($p < 1 \times 10^{-5}, 1 \times 10^{-4}, 2 \times 10^{-4}, 1 \times 10^{-3}$), corresponding to 71, 371, 623, and 1970 SNPs, respectively. Evaluation metrics are IBS, p_{out}, and d_{out}, with smaller values corresponding to better prediction performance

(a)							
Scenario	**APOE***	**PRS***	**icenReg**	**Bayesian**	**ALASSO**	**ICcforest**	**NN-IC**
	IBS (mean(sd))						
1	0.085(0.011)	0.084(0.012)	0.068(0.016)	0.064(0.005).	0.068(0.011)	0.081(0.013)	**0.065(0.002)**
2	0.085(0.011)	0.084(0.012)	0.064(0.018)	—	0.056(0.013)	0.082(0.013)	**0.053(0.006)**
3	0.085(0.011)	0.084(0.012)	—	—	—	0.082(0.012)	**0.050(0.006)**
4	0.085(0.011)	0.084(0.012)	—	—	—	0.083(0.013)	**0.040(0.007)**
	p_{out} (mean(sd))						
1	0.41(0.04)	0.41(0.04)	0.32(0.04)	0.32(0.05).	0.33(0.05)	0.42(0.05)	**0.33(0.04)**
2	0.41(0.04)	0.41(0.04)	0.44(0.04)	—	0.32(0.06)	0.42(0.05)	**0.30(0.05)**
3	0.41(0.04)	0.41(0.04)	—	—	—	0.42(0.05)	**0.32(0.04)**
4	0.41(0.04)	0.41(0.04)	—	—	—	0.43(0.05)	**0.28(0.03)**
	d_{out} (mean(sd))						
1	8.3(0.9)	8.3(0.9)	7.3(0.8)	7.5(0.7).	7.1(0.6)	7.8(1.2)	**7.8(0.6)**
2	8.3(0.9)	8.3(0.9)	12.8(0.6)	—	6.1(0.9)	8.0(1.3)	**6.7(1.0)**
3	8.3(0.9)	8.3(0.9)	—	—	—	8.0(1.2)	**6.2(1.1)**
4	8.3(0.9)	8.3(0.9)	—	—	—	8.1(1.3)	**5.8(0.8)**
(b)							
Scenario	**APOE***	**PRS***	**icenReg**	**Bayesian**	**ALASSO**	**ICcforest**	**NN-IC**
	IBS						
1	0.083	0.082	0.091	0.089	0.085	0.083	**0.084**
2	0.083	0.082	0.112	—	0.079	0.085	**0.073**
3	0.083	0.082	—	—	—	0.086	**0.070**
4	0.083	0.082	—	—	—	0.086	**0.069**
	p_{out}						
1	0.41	0.42	0.42	0.42	0.43	0.43	**0.46**
2	0.41	0.42	0.48	—	0.43	0.44	**0.46**
3	0.41	0.42	—	—	—	0.43	**0.45**
4	0.41	0.42	—	—	—	0.44	**0.45**
	d_{out}						
1	7.6	7.7	9.6	10.0	8.5	7.8	**8.9**
2	7.6	7.7	10.0	—	8.7	8.1	**7.2**
3	7.6	7.7	—	—	—	8.5	**7.2**
4	7.6	7.7	—	—	—	8.3	**6.6**

Note *The APOE and PRS models are invariant to the choice of p-value thresholds.

within each predictor), suggesting heterogeneity in the population.

We successfully identify two distinct subgroups with differential progression profiles using NN-IC in Data2. Specifically, we perform the Gaussian mixture model on the predicted prognostic index $\hat{g}$ (output from the neural network in NN-IC), as illustrated in the histogram of Figure 2A. The corresponding plot on the Turnbull estimates (Turnbull, 1976) of progression-free probabilities indicates significantly different progression profiles between the two subgroups (namely, the low-risk and high-risk subgroups), with $p = 7.4 \times 10^{-5}$ based on the log-rank test in interval-censored data (Finkelstein, 1986). Moreover, no subgroups can be detected based on the predicted log-likelihood values from the ICcforest model, as illustrated in Figure 2B. Overall, NN-IC's accurate prediction performance, the individualized predictor importance measures, and the identified risk subgroups provide valuable insights for early prevention and clinical management of AD.

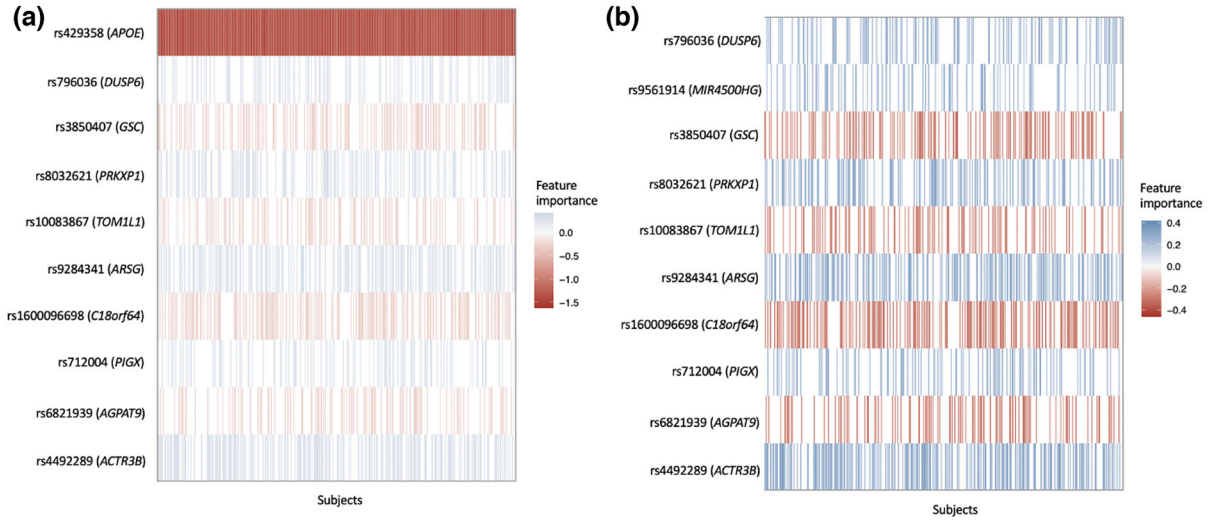

FIGURE 1 The representation of individualized importance measures for the top predictors with (A) and without (B) *APOE* in the external set of Data2 using the LIME method. Each row represents one predictor and each vertical column represents one subject. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

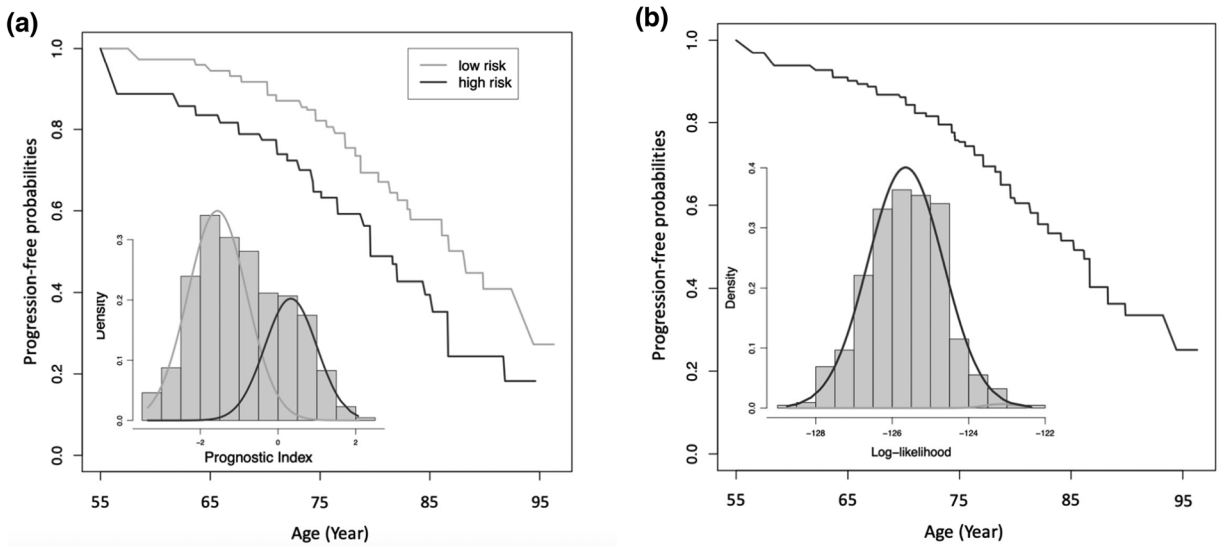

FIGURE 2 The Turnbull estimators of the progression-free probabilities of AD for the identified subgroups under NN-IC (A) and ICcforest (B) in the external set of Data2. The histograms show the predicted $\hat{g}$ values for NN-IC (A) and the predicted log-likelihood values for ICcforest (B), with subgroups identified by the Gaussian mixture model. Two subgroups are identified by NN-IC (log-rank test $p = 7.4 \times 10^{-5}$), and no subgroup is identified by ICcforest

7 | DISCUSSION AND CONCLUSION

Our work is the first study on AD prediction that appropriately handles interval-censoring and fully utilizes the wealthy GWAS predictors in the complete ADNI data. Our AD prediction model predicts risks of developing AD starting from an early age, enabling clinicians and patients to improve their quality of life before and during disease progression. Our prediction model achieves higher prediction accuracy than existing methods for interval-censored data and identifies high- and low-risk subgroups that facilitate early prevention.

Our work has two major contributions. First, we report the first GWAS that investigates genetic variants associated with AD development (i.e., time-to-AD) by utilizing the full ADNI genetic data. To deal with interval censoring and left-truncation issues in ADNI, we propose a semiparametric transformation model with a sieve

estimation approach. A computationally efficient score test is developed to enable the fast screening of millions of SNPs. We successfully identify multiple novel SNPs, which may advance the understanding of biological mechanisms underlying AD development. Second, our NN-IC method is the first neural network method for interval-censored data that employs a flexible semiparametric loss function. NN-IC incorporates and optimizes two neural networks simultaneously: a BPNet to estimate the infinite-dimensional baseline cumulative hazard function and another neural network to estimate the high-dimensional covariate effects. In particular, BPNet naturally satisfies the nonnegative and nondecreasing property of the cumulative hazard function by implementing Bernstein polynomials into the neural network. Overall, NN-IC achieves better accuracy than several existing methods in simulated and ADNI datasets. It can be readily applied to other progressive disorders where the event of interest is intermittently assessed.

Many machine/deep-learning methods have been developed for disease progression prediction. In particular, for the GWAS data, recent studies report that the neural network shows advantages in terms of prediction accuracy compared with other machine learning methods in right-censored data (Sun et al., 2020; Yan et al., 2021). To evaluate the prediction performance of neural networks in interval-censored data, we analyze two additional benchmark datasets. One is from the Age-related Eye Disease Study (AREDS) containing 7803 observations and 666 genetic variants. The event of interest is time-to-late-AMD (age-related macular degeneration), which is interval-censored due to intermittent assessments. Sun et al. (2020) used AREDS to predict AMD progression but imputed the interval-censored data into right-censored data. We split AREDS into training and test datasets at a ratio of 9:1 and report results in the test dataset. As shown in Web Table 4a, NN-IC exhibits better prediction performance than ICcforest, ALASSO, and icenReg. The second dataset is from the Tandmobiel study containing 61,267 observations and 49 clinical predictors, most of which are binary. The event of interest is time-to-tooth-emergence in children. The event is interval-censored due to annual dental examinations. Yao et al. (2021) used this dataset to evaluate the performance of ICcforest. We split the data into training and test datasets and report results in the test dataset. As shown in Web Table 4b, NN-IC and ICcforest exhibit similar prediction performance. In summary, NN-IC gives better prediction accuracy than other machine learning methods in the presence of high-dimensional genetic predictors.

One limitation of our work is that NN-IC involves tuning multiple hyperparameters. Based on our experience, we could start from a single hidden layer with 50 nodes and look for the other tuning parameters. Another limitation is that our AD prediction model uses only Caucasian individuals which constitutes 93% of all ADNI individuals. Because the mechanisms of AD development are reported to be different between Caucasian and non-Caucasian individuals (Morris et al., 2019), separate prediction models are needed for the two populations. It is desirable to develop a new prediction model based on non-Caucasian individuals in the future.

There are multiple directions to improve NN-IC in AD prediction. For example, we use only demographical and genetic predictors available at the baseline. The prediction accuracy could be further enhanced by incorporating time-dependent predictors such as longitudinally measured cognitive scores (Li & Luo, 2019; Jiang et al., 2021). Moreover, one may employ convoluted neural networks to exact useful features from medical images (i.e., structure MRI) and build a comprehensive prediction model using clinical, genetic, and imaging information.

ACKNOWLEDGMENTS

The authors thank the editor, associate editor, and referees for their insightful comments and suggestions that have led to a significant improvement of this paper. The authors acknowledge the support (to TS) from the National Natural Science Foundation of China (72101261, 72141306), National Bureau of Statistics of China (2021LZ18), the Ministry of Education of China (20JZD023), the Fund for Building World-class Universities (Disciplines) of Renmin University of China (KYGJC2021014), and Public Health & Disease Control and Prevention, Major Innovation & Planning Interdisciplinary Platform for the "Double-First Class" Initiative of Renmin University of China. Data used in the preparation of this article are obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. The public computing cloud from the Renmin University of China was used to perform the simulation and data analysis.

DATA AVAILABILITY STATEMENT

The data that support the findings of this paper are openly available from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu).

ORCID

Tao Sun https://orcid.org/0000-0003-4447-3005
Ying Ding https://orcid.org/0000-0003-1352-1000

REFERENCES

Addona, V. & Wolfson, D.B. (2006) A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up. *Lifetime Data Analysis*, 12(3), 267–284.

Alioum, A. & Commenges, D. (1996) A proportional hazards model for arbitrarily censored and truncated data. *Biometrics*, 52, 512–524.

Anderson-Bergman, C. (2017) icenReg: regression models for interval censored data in R. *Journal of Statistical Software*, 81(12), 1–23.

Bello, G.A., Dawes, T.J., Duan, J., Biffi, C., De Marvao, A., Howard, L.S., et al. (2019) Deep-learning cardiac motion analysis for human survival prediction. *Nature Machine Intelligence*, 1(2), 95–104.

Cheung, C.T., Singh, R., Kalra, R.S., Kaul, S.C. & Wadhwa, R. (2014) Collaborator of ARF (CARF) regulates proliferative fate of human cells by dose-dependent regulation of DNA damage signaling. *Journal of Biological Chemistry*, 289(26), 18258–18269.

Cox, D.R. & Hinkley, D.V. (1979) *Theoretical statistics*. Boca Raton, FL: CRC Press.

Desikan, R.S., Fan, C.C., Wang, Y., Schork, A.J., Cabral, H.J., Cupples, L.A., et al. (2017) Genetic assessment of age-associated Alzheimer disease risk: development and validation of a polygenic hazard score. *PLoS Medicine*, 14(3), e1002258.

Escott-Price, V., Shoai, M., Pither, R., Williams, J. & Hardy, J. (2017) Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. *Neurobiology of Aging*, 49, 214–e7.

Faraggi, D. & Simon, R. (1995) A neural network model for survival data. *Statistics in Medicine*, 14(1), 73–82.

Finkelstein, D.M. (1986) A proportional hazards model for interval-censored failure time data. *Biometrics*, 42, 845–854.

Gao, F. & Chan, K. C.G. (2019) Semiparametric regression analysis of length-biased interval-censored data. *Biometrics*, 75(1), 121–132.

Gatz, M., Reynolds, C.A., Fratiglioni, L., Johansson, B., Mortimer, J.A., Berg, S., et al. (2006) Role of genes and environments for explaining Alzheimer's disease. *Archives of General Psychiatry*, 63(2), 168–174.

Hornik, K., Stinchcombe, M. & White, H. (1989) Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366.

Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., et al. (2019) Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics*, 51(3), 404–413.

Jiang, S., Xie, Y. & Colditz, G.A. (2021) Functional ensemble survival tree: dynamic prediction of Alzheimer's disease progression accommodating multiple time-varying covariates. *Journal of the Royal Statistical Society: Series C*, 70(1), 66–79.

Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T. & Kluger, Y. (2018) DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Medical Research Methodology*, 18(1), 24.

Kawamata, J. & Shimohama, S. (2002) Association of novel and established polymorphisms in neuronal nicotinic acetylcholine receptors with sporadic Alzheimer's disease. *Journal of Alzheimer's Disease*, 4(2), 71–76.

Kong, D., Ibrahim, J.G., Lee, E. & Zhu, H. (2018) FLCRM: functional linear Cox regression model. *Biometrics*, 74(1), 109–117.

Lee, M., Gouskova, N.A., Feuer, E.J. & Fine, J.P. (2017) On the choice of time scales in competing risks predictions. *Biostatistics*, 18(1), 15–31.

Li, C., Pak, D. & Todem, D. (2020) Adaptive lasso for the Cox regression with interval censored and possibly left truncated data. *Statistical Methods in Medical Research*, 29(4), 1243–1255.

Li, H., Habes, M., Wolk, D.A. & Fan, Y. (2019) A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimer's & Dementia*, 15(8), 1059–1070.

Li, K. & Luo, S. (2019) Dynamic prediction of Alzheimer's disease progression using features of multiple longitudinal outcomes and time-to-event data. *Statistics in Medicine*, 38(24), 4804–4818.

Lin, J., Li, K. & Luo, S. (2021) Functional survival forests for multivariate longitudinal outcomes: dynamic prediction of Alzheimer's disease progression. *Statistical Methods in Medical Research*, 30(1), 99–111.

Lin, X., Cai, B., Wang, L. & Zhang, Z. (2015) A Bayesian proportional hazards model for general interval-censored data. *Lifetime Data Analysis*, 21(3), 470–490.

Morris, J.C., Schindler, S.E., McCue, L.M., Moulder, K.L., Benzinger, T.L., Cruchaga, C., et al. (2019) Assessment of racial disparities in biomarkers for Alzheimer's disease. *JAMA Neurology*, 76(3), 264–273.

Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C., Jagust, W., et al. (2005) The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics of North America*, 15(4), 869.

Nakagawa, T., Ishida, M., Naito, J., Nagai, A., Yamaguchi, S. & Onoda, K. (2020) Prediction of conversion to Alzheimer's disease using deep survival analysis of MRI images. *Brain Communications*, 2(1), fcaa057.

Pan, W. & Chappell, R. (2002) Estimation in the Cox proportional hazards model with left-truncated and interval-censored data. *Biometrics*, 58(1), 64–70.

Petersen, R.C. (2009) Early diagnosis of Alzheimer's disease: Is MCI too late?. *Current Alzheimer Research*, 6(4), 324–330.

Privé, F., Aschard, H., Ziyatdinov, A. & Blum, M.G. (2018) Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. *Bioinformatics*, 34(16), 2781–2787.

Reitz, C., Rogaeva, E. & Beecham, G.W. (2020) Late-onset vs nonmendelian early-onset Alzheimer disease: a distinction without a difference?. *Neurology Genetics*, 6(5), 1–9.

Ribeiro, M.T., Singh, S. & Guestrin, C. (2016) Why should I trust you? Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* . New York: ACM, pp. 1135–1144.

Shen, P.-S. (2014) Proportional hazards regression with interval-censored and left-truncated data. *Journal of Statistical Computation and Simulation*, 84(2), 264–272.

Shen, P.-S., Chen, H.-J., Pan, W.-H. & Chen, C.-M. (2019) Semiparametric regression analysis for left-truncated and interval-censored data without or with a cure fraction. *Computational Statistics & Data Analysis*, 140, 74–87.

Sun, T. & Ding, Y. (2021) Copula-based semiparametric regression method for bivariate data under general interval censoring. *Biostatistics*, 22(2), 315–330.

Sun, T., Liu, Y., Cook, R.J., Chen, W. & Ding, Y. (2019) Copula-based score test for bivariate time-to-event data, with application to a genetic study of AMD progression. *Lifetime Data Analysis*, 25(3), 546–568.

Sun, T., Wei, Y., Chen, W. & Ding, Y. (2020) Genome-wide association study-based deep learning for survival prediction. *Statistics in Medicine*, 39(30), 4605–4620.

Tsouprou, S., Putter, H. & Fiocco, M. (2015) *Measures of discrimination and predictive accuracy for interval censored survival data*. (Doctoral dissertation, Master's Thesis]: Leiden University. http://www.math.leidenuniv.nl/scripties/MasterTsouprou.pdf)

Turnbull, B.W. (1976) The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society Series B*, 38(3), 290–295.

Wang, P., Li, D. & Sun, J. (2021) A pairwise pseudo-likelihood approach for left-truncated and interval-censored data under the Cox model. *Biometrics*, 77(4), 1303–1314.

Wu, Q., Zhao, H., Zhu, L. & Sun, J. (2020) Variable selection for high-dimensional partly linear additive Cox model with application to Alzheimer's disease. *Statistics in Medicine*, 39(23), 3120–3134.

Wu, Y., Wang, X., Lin, J., Jia, B. & Owzar, K. (2020) Predictive accuracy of markers or risk scores for interval censored survival data. *Statistics in Medicine*, 39(18), 2437–2446.

Yan, Q., Jiang, Y., Huang, H., Swaroop, A., Chew, E.Y., Weeks, D.E., et al. (2021) Genome-wide association studies-based machine learning for prediction of age-related macular degeneration risk. *Translational Vision Science & Technology*, 10(2), 29–29.

Yao, W., Frydman, H. & Simonoff, J.S. (2021) An ensemble method for interval-censored time-to-event data. *Biostatistics*, 22(1), 198–213.

Zhou, Q., Hu, T. & Sun, J. (2017) A sieve semiparametric maximum likelihood approach for regression analysis of bivariate interval-censored failure time data. *Journal of the American Statistical Association*, 112(518), 664–672.

SUPPORTING INFORMATION

Web Tables and Figures referenced in Sections 5, 6, and 7 are available with this paper at the Biometrics website on Wiley Online Library. The R-code is provided on Wiley and GitHub (suntaojj/NN-IC).

How to cite this article: Sun, T. & Ding, Y. (2022) Neural network on interval-censored data with application to the prediction of Alzheimer's disease. *Biometrics*, 1–14. https://doi.org/10.1111/biom.13734