



## OPEN ACCESS

## EDITED BY

Mohammad Taheri,  
University Hospital Jena, Germany

## REVIEWED BY

Hikoaki Fukaura,  
Saitama Medical University, Japan  
Yusei Miyazaki,  
National Hospital Organization, Japan  
Yaqing Shu,  
Third Affiliated Hospital of Sun Yat-sen  
University, China

## \*CORRESPONDENCE

Chao Quan  
chao\_quan@fudan.edu.cn

## SPECIALTY SECTION

This article was submitted to  
Multiple Sclerosis and  
Neuroimmunology,  
a section of the journal  
Frontiers in Neurology

RECEIVED 19 May 2022

ACCEPTED 11 July 2022

PUBLISHED 05 August 2022

## CITATION

Wang L, Du L, Li Q, Li F, Wang B,  
Zhao Y, Meng Q, Li W, Pan J, Xia J,  
Wu S, Yang J, Li H, Ma J, ZhangBao J,  
Huang W, Chang X, Tan H, Yu J,  
Zhou L, Lu C, Wang M, Dong Q, Lu J,  
Zhao C and Quan C (2022) Deep  
learning-based relapse prediction of  
neuromyelitis optica spectrum  
disorder with anti-aquaporin-4  
antibody. *Front. Neurol.* 13:947974.  
doi: 10.3389/fneur.2022.947974

## COPYRIGHT

© 2022 Wang, Du, Li, Li, Wang, Zhao,  
Meng, Li, Pan, Xia, Wu, Yang, Li, Ma,  
ZhangBao, Huang, Chang, Tan, Yu,  
Zhou, Lu, Wang, Dong, Lu, Zhao and  
Quan. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Deep learning-based relapse prediction of neuromyelitis optica spectrum disorder with anti-aquaporin-4 antibody

Liang Wang<sup>1,2</sup>, Lei Du<sup>3</sup>, Qinying Li<sup>4</sup>, Fang Li<sup>2,5</sup>, Bei Wang<sup>6</sup>, Yuanqi Zhao<sup>7</sup>, Qiang Meng<sup>8</sup>, Wenyu Li<sup>9</sup>, Juyuan Pan<sup>10</sup>, Junhui Xia<sup>10</sup>, Shitao Wu<sup>11</sup>, Jie Yang<sup>12</sup>, Heng Li<sup>13</sup>, Jianhua Ma<sup>3</sup>, Jingzi ZhangBao<sup>1,2</sup>, Wenjuan Huang<sup>1,2</sup>, Xuechun Chang<sup>1,2</sup>, Hongmei Tan<sup>1,2</sup>, Jian Yu<sup>14</sup>, Lei Zhou<sup>1,2</sup>, Chuanzhen Lu<sup>1,2</sup>, Min Wang<sup>14</sup>, Qiang Dong<sup>1,2</sup>, Jiahong Lu<sup>1,2</sup>, Chongbo Zhao<sup>1,2</sup> and Chao Quan<sup>1,2\*</sup> on behalf of the Pan-Yangtze River Delta Alliance for Demyelinating Disease

<sup>1</sup>Department of Neurology, Huashan Rare Disease Center, Huashan Hospital, Shanghai Medical College, Fudan University, Shanghai, China, <sup>2</sup>National Center for Neurological Disorders (NCND), Shanghai, China, <sup>3</sup>Department of Neurology, The First Affiliated Hospital of Xinjiang Medical University, Urumqi, Xinjiang Uygur Autonomous Region, China, <sup>4</sup>Department of Rehabilitation Medicine, Jing'an District Centre Hospital of Shanghai, Fudan University, Shanghai, China, <sup>5</sup>Department of Rehabilitation Medicine, Huashan Hospital, Shanghai Medical College, Fudan University, Shanghai, China, <sup>6</sup>Department of Neurology, Jing'an District Centre Hospital of Shanghai, Fudan University, Shanghai, China, <sup>7</sup>Department of Neurology, The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, China, <sup>8</sup>Department of Neurology, The First People's Hospital of Yunnan Province, Kunming, China, <sup>9</sup>Department of Neurology, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, China, <sup>10</sup>Department of Neurology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, <sup>11</sup>Department of Neurology, The Fifth Affiliated Hospital of Zhengzhou University, Zhengzhou, China, <sup>12</sup>Department of Neurology, Wuhan No.1 Hospital, Wuhan, China, <sup>13</sup>Department of Neurology, Central Hospital Affiliated to Shandong First Medical University, Jinan, China, <sup>14</sup>Department of Ophthalmology and Vision Science, Eye and ENT Hospital, Shanghai Medical College, Fudan University, Shanghai, China

**Objective:** We previously identified the independent predictors of recurrent relapse in neuromyelitis optica spectrum disorder (NMOSD) with anti-aquaporin-4 antibody (AQP4-ab) and designed a nomogram to estimate the 1- and 2-year relapse-free probability, using the Cox proportional hazard (Cox-PH) model, assuming that the risk of relapse had a linear correlation with clinical variables. However, whether the linear assumption fits real disease tragedy is unknown. We aimed to employ deep learning and machine learning to develop a novel prediction model of relapse in patients with NMOSD and compare the performance with the conventional Cox-PH model.

**Methods:** This retrospective cohort study included patients with NMOSD with AQP4-ab in 10 study centers. In this study, 1,135 treatment episodes from 358 patients in Huashan Hospital were employed as the training set while 213 treatment episodes from 92 patients in nine other research centers as the validation set. We compared five models with added variables of gender,

AQP4-ab titer, previous attack under the same therapy, EDSS score at treatment initiation, maintenance therapy, age at treatment initiation, disease duration, the phenotype of the most recent attack, and annualized relapse rate (ARR) of the most recent year by concordance index (C-index): conventional Cox-PH, random survival forest (RSF), LogisticHazard, DeepHit, and DeepSurv.

**Results:** When including all variables, RSF outperformed the C-index in the training set (0.739), followed by DeepHit (0.737), LogisticHazard (0.722), DeepSurv (0.698), and Cox-PH (0.679) models. As for the validation set, the C-index of LogisticHazard outperformed the other models (0.718), followed by DeepHit (0.704), DeepSurv (0.698), RSF (0.685), and Cox-PH (0.651) models. Maintenance therapy was calculated to be the most important variable for relapse prediction.

**Conclusion:** This study confirmed the superiority of deep learning to design a prediction model of relapse in patients with AQP4-ab-positive NMOSD, with the LogisticHazard model showing the best predictive power in validation.

#### KEYWORDS

neuromyelitis optica spectrum disorder, anti-aquaporin-4 antibody, machine learning, deep learning, relapse prediction

## Introduction

Neuromyelitis optica spectrum disorder (NMOSD) is an inflammatory disease of the central nervous system, mainly manifesting as relapsing optic neuritis (ON) and transverse myelitis (TM), resulting in visual and motor disability (1). The majority of patients with NMOSD harbor pathogenic autoantibodies targeting aquaporin 4 (AQP4) water channels in the serum. Disability accumulation in NMOSD is relapse-dependent, thus preventing or delaying relapses is the primary goal in NMOSD management (2). Traditional first-line immunosuppressants used as maintenance therapy in NMOSD include azathioprine (AZA), mycophenolate mofetil (MMF), and rituximab (RTX), while novel monoclonal antibodies such as satralizumab, eculizumab, and inebilizumab have exhibited powerful efficacy in controlling relapses recently (3–7). Recognizing the risk factors of relapse and establishing a suitable prediction model are of utmost importance to inform individualized therapy for NMOSD.

Survival analysis (also called time-to-event analysis) has been widely used to estimate the probability of prognostic outcomes such as death or disease recurrence. The Cox proportional hazard (Cox-PH) model is the most well-known approach for determining the association between a predictive variable of clinical characteristics and the risk of an event such as death (8). Nomogram is a feasible prediction model using risk factors to estimate the probability of a certain event. We previously identified the independent predictors of recurrent relapse in NMOSD, including gender, anti-AQP4 antibody (AQP4-ab) titer, previous attack under the same therapy, EDSS

score at treatment initiation, and maintenance therapy, and designed a nomogram to estimate the 1- and 2-year relapse-free probability (9). However, this model was based on the assumption that the risk of a certain event had a linear combination with the variables, which could be too simplistic to fit the actual disease trajectory.

Machine learning is one branch of artificial intelligence and has a wide range of applications, such as predicting carcinoma development and cardiovascular events (10, 11). Compared with the conventional Cox-PH model, machine learning has several advantages, including the ability to continually incorporate new data to optimize algorithms and identify clinically important risks with some marginal variables (12). Deep learning, a novel form of machine learning, employs improved computer power and big data to outperform other machine-learning techniques (13). It has also been used in imaging diagnosis, disease staging, and prognosis and has proved to improve outcome prediction (14–16). The aim of the current investigation was to employ deep learning and machine learning to design a novel prediction model of relapse in patients with NMOSD with AQP4-ab and compare its performance with that of the conventional Cox-PH model.

## Methods

### Study design and participants

This retrospective cohort study included a cohort of patients with NMOSD with AQP4-ab, based on which an outcome

prediction model of NMOSD relapse under certain treatment was published by our group (9). We divided NMOSD treatments during follow-up into censored (relapse-free) or failed (disease relapse) treatment episodes. The exclusion criteria were as follows: (1) less than 15 episodes of a certain treatment; (2) lasting for <3 months; (3) without definite start or stop dates; and (4) receiving double/overlapping medications. The effectiveness duration from the last administration time of each medication referred to the study by Stellmann et al. (17).

Among them, 1,135 treatment episodes from 358 patients in Huashan Hospital were employed as the training set while 213 treatment episodes from 92 patients in nine other research centers as the validation set. Each treatment episode was regarded as one independent individual. Demographic and clinical characteristics were collected, including gender, age at treatment initiation, disease duration, AQP4-ab titer, annualized relapse rate (ARR) of the most recent year, Expanded Disability Status Scale (EDSS) score at treatment initiation, the previous attack under same therapy, phenotype of the most recent attack, and maintenance therapy.

All patients from the above research centers received serum AQP4-ab and MOG-ab detection using fixed cell-based indirect immune-fluorescence tests. HEK293 cells transfected with AQP4 M1 isoform or full-length human MOG were employed. AQP4-ab titer  $\geq 1:100$  was identified as a high level. MOG-ab was not found in all AQP4 ab-positive patients.

## Statistical analyses

Data analysis and graphing were performed using SPSS version 22.0 (SPSS Inc., Chicago IL, USA), GraphPad Prism 6 (GraphPad Software Inc., La Jolla CA, USA), the rms, survival, and survminer packages of R (version 4.1.3, <http://www.r-project.org/>). Discrete variables were expressed as count and percentage, while continuous variables were presented as means  $\pm$  1 standard deviation or medians with four quantile ranges. Kaplan–Meier survival analysis was used for the relapse curves of the training and validation sets, as well as the training set with various maintenance therapies. Harrell's concordance index (C-index) was deemed as the most suitable and accurate approach for estimating prediction error, which measured the concordance between the predicted and actual probability (18). A Harrell's C-index of 0.5 reveals no predictive discrimination, >0.7 reveals a good model, and > 0.8 reveals a strong model (19). We compared five models with added variables of gender, AQP4-ab titer, previous attack under the same therapy, EDSS score at treatment initiation, maintenance therapy, age at treatment initiation, disease duration, the phenotype of the most recent attack, and ARR of the most recent year by C-index: conventional Cox-PH, random survival forest (RSF),

LogisticHazard, DeepHit, and DeepSurv. Statistical significance was set at  $p < 0.05$ .

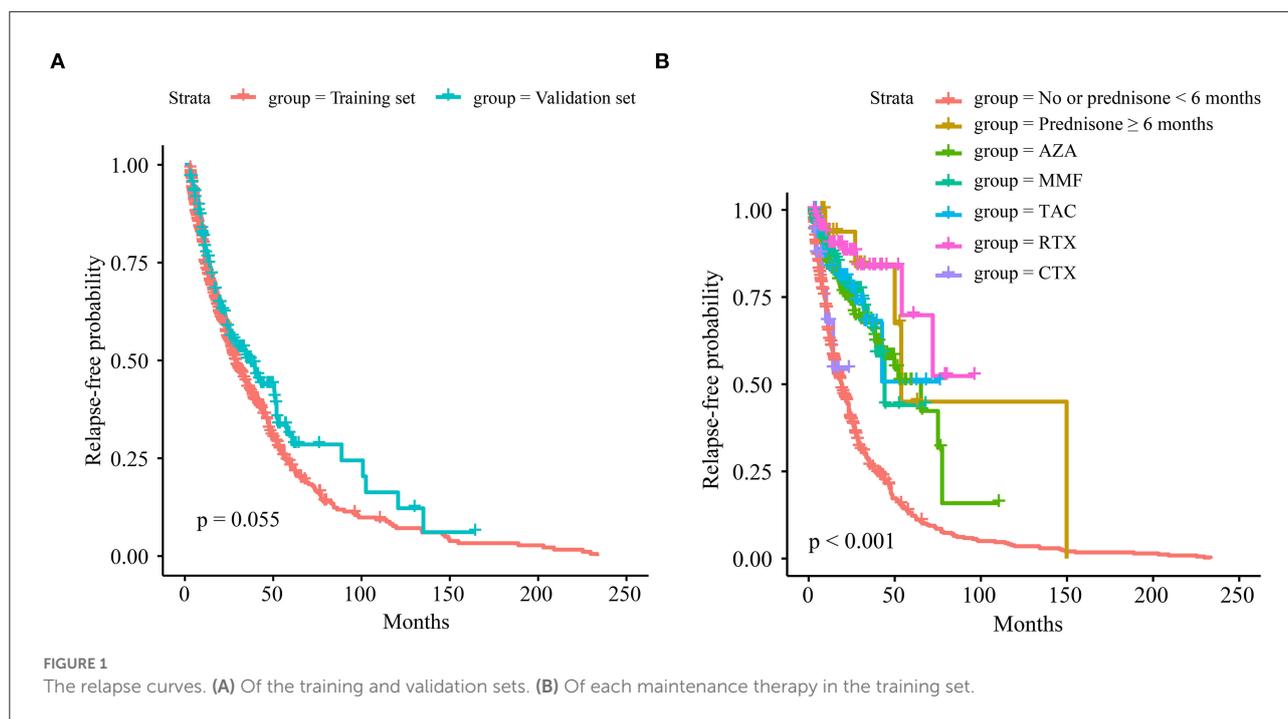
## Random survival forest

Building the RSF model and graphing were conducted with RandomForestSRC, ggRandomForests, and ggplot2 packages of R (version 4.1.3, <http://www.r-project.org/>). It is a nonparametric model that builds hundreds of trees and outputs results in the form of voting (20). This model reduces variance and bias by employing all variables collected and automatically evaluating complex interactions and

TABLE 1 The demographic and clinical characteristics of AQP4-ab positive NMOSD patients in the training and validation set.

Patient characteristics	Training set (n = 1,135)	Validation set (n = 213)
Female gender, n (%)	1,055 (93.0)	198 (93.0)
Age at treatment initiation, years	35.7 (26.2–47.8)	47.5 (34–54.9)
Disease duration, months	18.3 (1.1–49.7)	9.7 (1–29.3)
High AQP4-ab titer ( $\geq 1:100$ ), n (%)	643 (56.7)	158 (74.2)
ARR of the most recent year	1 (1–1)	1 (1–2)
EDSS score at treatment initiation	2 (1–3)	2 (1–3)
Previous attack under same therapy, n (%)	419 (36.9)	71 (33.3)
<b>Most recent attack, n (%)</b>		
ON	422 (37.2)	40 (18.8)
TM	432 (38.1)	121 (56.8)
Brainstem/cerebral	117 (10.3)	8 (3.8)
Mixed	164 (14.4)	44 (20.7)
<b>Maintenance therapy, n (%)</b>		
No or prednisone <6 months	612 (53.9)	110 (51.6)
Prednisone ( $\geq 6$ months)	19 (1.7)	9 (4.2)
AZA	191 (16.8)	42 (19.7)
MMF	164 (14.4)	27 (12.7)
TAC	46 (4.1)	1 (0.5)
RTX	86 (7.6)	16 (7.5)
CTX	17 (1.5)	8 (3.8)

AQP4-ab, aquaporin-4 antibody; ARR, annualized relapse rate; EDSS, Expanded Disability Status Scale; ON, optic neuritis; TM, transverse myelitis; AZA, azathioprine; MMF, mycophenolate mofetil; TAC, tacrolimus; RTX, rituximab; CTX, cyclophosphamide.



nonlinear effects (21). The C-index equals the sum of consistent logarithms divided by the total number of data pairs, with a prediction error rate of 1 minus the C-index. The importance of variables was judged by the minimal depth and variable importance (VIMP) method. Higher VIMP or lower minimal depth contributed more to predicting accuracy (22).

## Deep learning

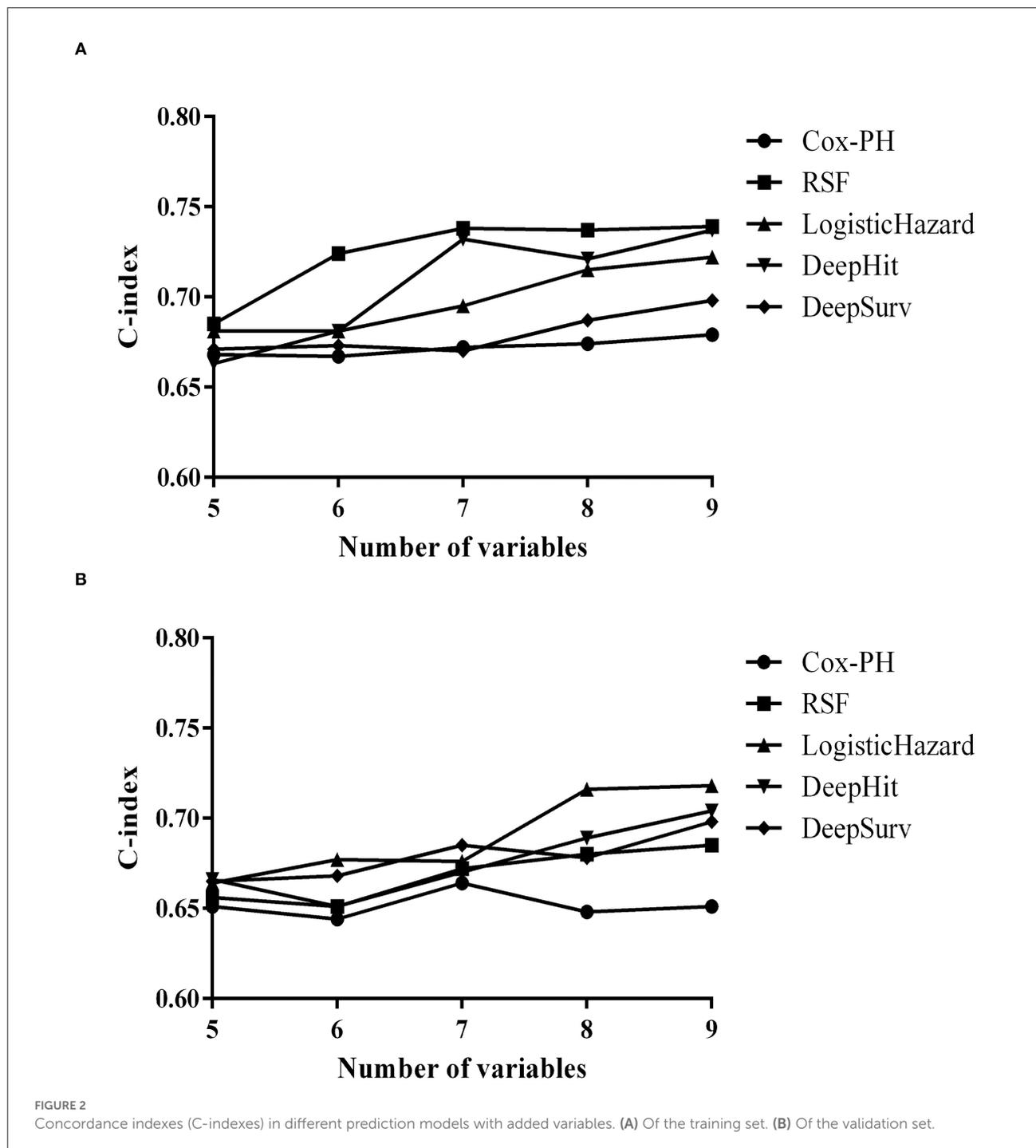
We used three methods of deep learning, which were implemented in Pytorch with the Python package pycox (version 3.7.0, <https://www.python.org/>). The parameters in the model were manually optimized, with dropouts ranging from 0 to 1. LogisticHazard, also called Nnet-survival, parametrizes the discrete hazards and optimizes the survival likelihood (23). One of the interpolation schemes, called constant hazard interpolation (CHI), was deemed as the C-index. DeepHit is a deep neural network trained by using a loss function, which exploits both relative risks and survival times, whose form of the stochastic process and parameters are dependent on the variables (24). DeepSurv, a multilayer feed-forward neural network, outputs a negative log partial likelihood, which is parameterized by the weights of the network (25). The C-indexes and losses in the training and validation sets were calculated, respectively.

## Results

### Baseline characteristics

The demographic and clinical characteristics of patients with AQP4-ab positive NMOSD in the training and validation sets are demonstrated in Table 1. We included 1,135 counts from Huashan Hospital as the training set and 213 counts from other centers as the validation set. Female patients were the predominant constituent, while approximately one-third of them had the previous attack under the same therapy. ON and TM were the most common recent attacks, with the EDSS score at treatment initiation ranging from 0 to 8.5. More than one-half of the counts had no maintenance therapy or prednisone for <6 months, and more than 10% of the counts had AZA and MMF as their maintenance therapy, respectively.

The relapse curves of the training and validation sets are indicated in Figure 1A, while the difference was not statistically significant ( $p = 0.055$ ). The relapse curves of each maintenance therapy in the training set are shown in Figure 1B, and the difference among them was highly statistically significant ( $p < 0.001$ ). Pairwise comparisons of maintenance therapy using a log-rank test are exhibited in Supplementary Table S1. Among them, compared with no or prednisone <6 months, maintenance therapy of prednisone  $\geq$ 6 months, AZA, MMF, tacrolimus (TAC), or RTX could efficiently control relapse with high statistical significance (all



$p < 0.001$ ). Compared with CTX, prednisone  $\geq 6$  months, AZA, MMF, and RTX could reduce relapse with significance ( $p = 0.017, 0.026, 0.011,$  and  $0.002$ , respectively). RTX was associated with a statistically significant reduction in relapse when compared with AZA and MMF ( $p = 0.014$  and  $0.042$ , respectively).

## Overall comparison

Prediction models of relapse based on Cox-PH, RSF, LogisticHazard, DeepHit, or DeepSurv patterns were established with the training set and verified in the validation set. The overall comparison of C-indexes with different dropouts using

TABLE 2 Independent predictors of relapse with multivariate Cox-PH model.

Independent predictors	Model	
	Hazard ratio (95% CI)	p-Value
Female gender (Reference = male)	1.40 (0.98–1.99)	0.063
AQP4-ab titer (Reference = <1:100)	1.29 (1.09–1.53)	0.003**
Previous attack under same therapy (Reference = no)	1.26 (1.02–1.56)	0.033*
EDSS score at treatment initiation (Reference = <2.5)	0.90 (0.84–0.97)	0.003**
<b>Maintenance therapy (Reference = no or prednisone &lt;6 months)</b>		
Prednisone ( $\geq 6$ months)	0.28 (0.11–0.68)	0.005**
AZA	0.39 (0.29–0.53)	< 0.001***
MMF	0.33 (0.23–0.48)	< 0.001***
TAC	0.34 (0.19–0.61)	< 0.001***
RTX	0.18 (0.10–0.33)	< 0.001***
CTX	0.94 (0.38–2.29)	0.89
Age at treatment initiation, years	1.01 (1.00–1.01)	0.11
Disease duration, months	1.00 (1.00–1.00)	0.08
<b>Phenotype of the most recent attack (Reference = brainstem/cerebral)</b>		
ON	0.86 (0.65–1.14)	0.30
TM	0.94 (0.71–1.25)	0.69
Mixed	1.00 (0.72–1.39)	1.00
ARR of the most recent year	1.20 (1.02–1.41)	0.026*

AQP4-ab, aquaporin-4 antibody; EDSS, Expanded Disability Status Scale; AZA, azathioprine; MMF, mycophenolate mofetil; TAC, tacrolimus; RTX, rituximab; CTX, cyclophosphamide; ON, optic neuritis; TM, transverse myelitis; ARR, annualized relapse rate.

\* $p < 0.05$ .

\*\* $p < 0.01$ .

\*\*\* $p < 0.001$ .

deep learning models in the training set could be seen in [Supplementary Figure S1](#), and the dropout in each model was set at 0.2 according to the optimization results. C-indexes in different prediction models with added variables of the training and validation sets are exhibited in [Figures 2A,B](#).

Initiating with five, the variables were added up to nine incrementally. The first five variables were gender, AQP4-ab, previous attack under the same therapy, EDSS score at treatment initiation, and maintenance therapy, which were statistically significant in our previous study (9). Then, statistically insignificant but clinically important variables were successively included, from age at treatment start, duration of disease, and phenotype of the most recent attack, to ARR of the most recent year.

When including all variables, RSF outperformed the training set (0.739), followed by DeepHit (0.737), LogisticHazard (0.722), DeepSurv (0.698), and Cox-PH (0.679) models. As for the

validation set, LogisticHazard outperformed the other models (0.718), followed by DeepHit (0.704), DeepSurv (0.698), RSF (0.685), and Cox-PH (0.651) models.

## Cox-PH model

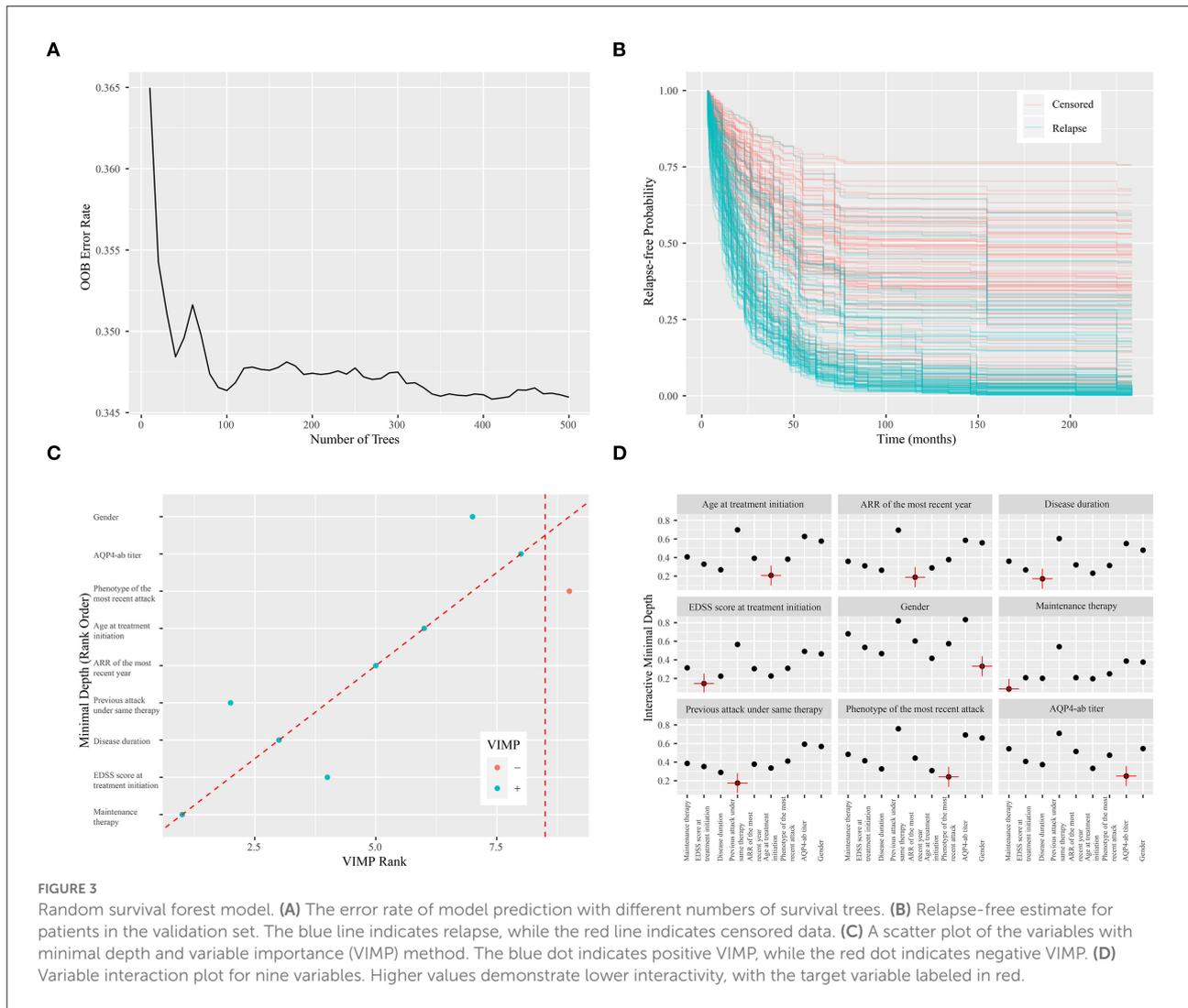
As the number of variables added up from 5 to 9, the C-index did not increase gradually as some were statistically insignificant variables. Finally, the C-index in the training set rose from 0.668 to 0.679 while in the validation set went up from 0.651 to 0.664 and fell back to 0.651. Independent predictors of relapse with the multivariate Cox-PH model are shown in [Table 2](#), in which AQP4-ab titer, previous attack under the same therapy, EDSS score at treatment initiation, maintenance therapy, and ARR of the most recent year were statistically significant relapse predictors ( $p = 0.003, 0.033, 0.003, <0.001, \text{ and } 0.026$ ).

## RSF model

There was a steady increase in the C-index as the number of variables built in the model expanded. Finally, the C-index in the training set increased from 0.685 to 0.739 while in the validation set from 0.656 to 0.685. The error rate of model prediction with different numbers of survival trees is shown in [Figure 3A](#). The model generated a total of 500 binary classification trees. It could be seen that the decreasing trend in the error rate has slowed down significantly. The error rates in the training and validation sets were 0.261 and 0.315, respectively. Relapse-free estimate for patients in the validation set is exhibited in [Figure 3B](#), with the blue line indicating relapse and the red line indicating censored data. Predictors calculated with minimal depth and VIMP from the RSF model are exhibited in [Supplementary Table S2](#). Considering lower minimal depth or higher VIMP contributed more to predicting accuracy, maintenance therapy was regarded as the paramount variable, with a minimal depth of 1.116 and VIMP of 0.213. A scatter plot of the variables with minimal depth and the VIMP method is shown in [Figure 3C](#) to provide a comprehensive view of variable rank. A variable interaction plot of minimal depth for nine variables is exhibited in [Figure 3D](#). Considering higher values equal lower interactivity, disease duration seems to have an association with other variables.

## LogisticHazard model

There was a steady increase in the C-index as the number of variables built in the model expanded. Finally, the C-index in the training set went up from 0.681 to 0.722, while in the validation set from 0.664 to 0.718, the highest among these models. The losses in the training and validation sets with the LogisticHazard model are shown in [Figure 4A](#).



## DeepHit model

As the number of variables added up from 5 to 9, the trend first increased and then fluctuated in the C-index, reaching 0.737 and 0.704 in the training and validation sets, respectively. The losses in the training and validation sets with the DeepHit model are presented in Figure 4B. The loss is the lowest among these deep learning models.

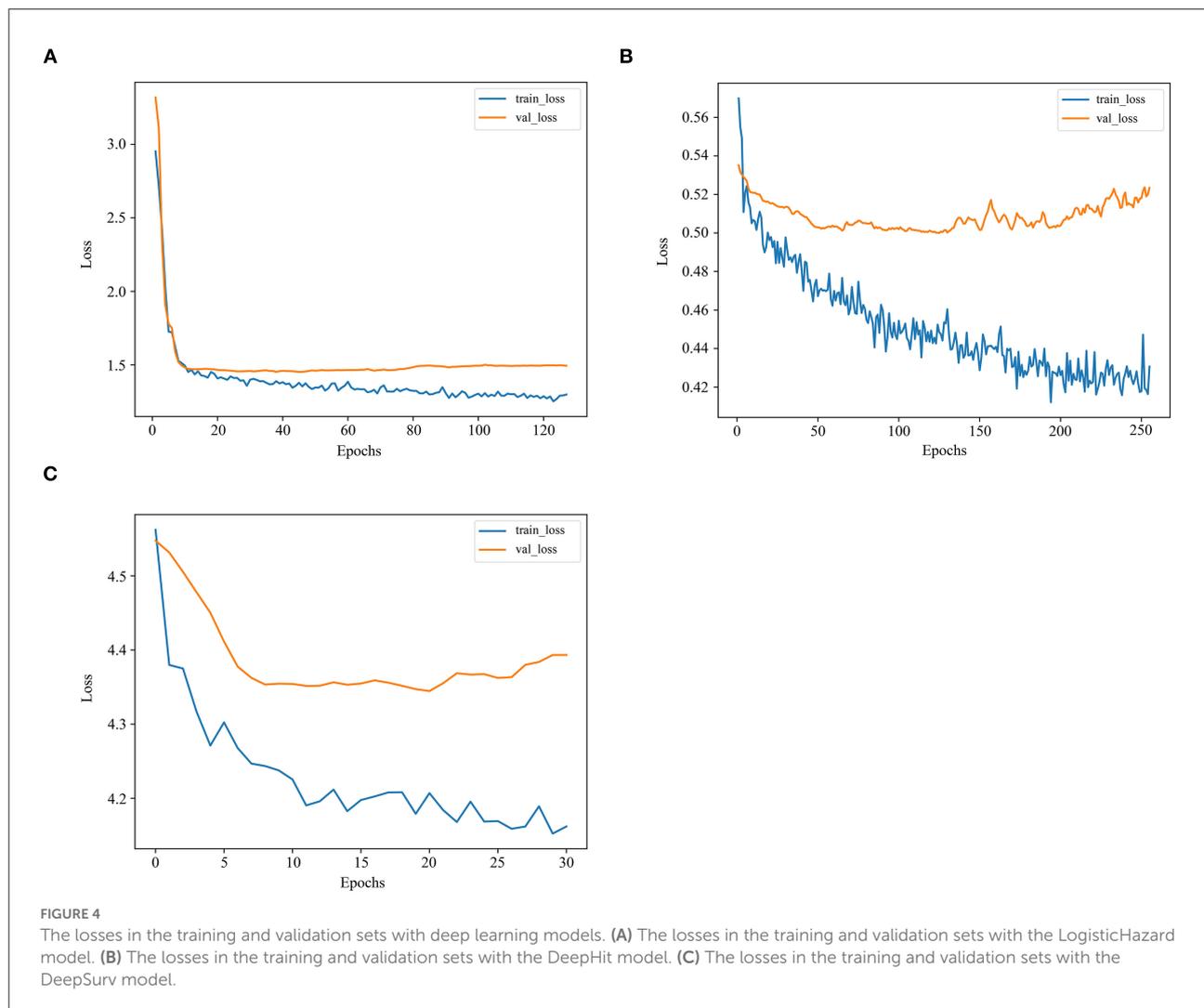
## DeepSurv model

As the number of variables added up from 5 to 9, the C-index first remained steady and then increased, reaching 0.698 and 0.698 in the training and validation sets, respectively. The losses in the training and validation sets with the DeepSurv model are exhibited in Figure 4C. The loss is

much higher in this model than in the other two deep learning models.

## Discussion

Previous studies regarding neurology have employed deep learning in stroke diagnosis and outcome prediction, as well as differentiating NMOSD from MS (26, 27). To the best of our knowledge, this was the first study employing artificial intelligence to design the prediction model of relapse in patients with NMOSD with AQP4-ab. It is beneficial for patients to stratify their risk of relapse and avoid ineffective or unnecessary treatment. The deep learning models demonstrated superior predictive power compared with the conventional CoxPH model, with the LogisticHazard model showing the best predictive power in validation.



The above prediction models included nine variables, namely, gender, AQP4-ab titer, previous attack under the same therapy, EDSS score at treatment initiation, maintenance therapy, age at treatment initiation, disease duration, the phenotype of the most recent attack, and ARR of the most recent year, all of which were common demographical and clinical characteristics. Maintenance therapy was calculated to be the most important variable, which was consistent with the clinical practice (17, 28, 29). Patients with NMOSD underwent more potent immunosuppressive treatment that had less rates of relapse, while those who received disease-modifying treatment had more rates of relapse (3, 28). It emphasized the importance of maintenance therapy as well as the correct differential diagnosis. The previous attack under the same therapy was another important variable, demonstrating that patients with repetitive attacks on certain treatment should convert to a more potent immunosuppressive drug, such as novel monoclonal antibodies such as satralizumab, eculizumab, and inebilizumab.

Disease duration, age, clinical manifestation, and ARR before treatment reached statistical significance in other studies and were regarded clinically important to be included in the prediction model (3, 17, 29).

The Cox-PH model is a classical approach to survival analysis and event prediction. However, the model is semiparametric, assuming that the risk of the event has a linear association with the variables. The advantage of the RSF model is that it is not constrained by the assumption of proportional hazard and log-linearity (21). Meanwhile, it could prevent the overfitting problem of its algorithm through two random sampling processes (30). The deep learning model could learn and infer high-order nonlinear combinations between patient outcomes and variables in a fully data-driven manner. Recently, it was demonstrated that deep neural network outperforms standard survival analysis, one of whose advantages is the ability to discern relationships without prior selection of features (25). In this study, the RSF

model fits best in the training set but not in the validation set, which may demonstrate some overfitting in this model. The LogisticHazard model performs best in the validation set with added variables, indicating its superior performance and improvements compared with the other models. The loss may partially explain the relative inferiority of the C-index in the DeepSurv model.

The primary limitation was the retrospective nature of this study, with some recall bias. First, the dose of maintenance therapies was varied and not standardized to the same level. Second, AQP4-ab titers were not examined at fixed timing after NMOSD onset. The black-box essence of the deep learning model limited its clinical utility. Future prospective and large-sample studies with advanced technology will further prompt to evolution and visualization of the model.

To conclude, this study confirmed the superiority of deep learning to design a prediction model of relapse in patients with AQP4-ab-positive NMOSD, with the LogisticHazard model showing the best predictive power in validation. Based on the above-optimized model, a personalized treatment recommender system could be developed to minimize the probability of relapse in the future.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by Medical Ethics Committee of Huashan Hospital. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

LW designed and conceptualized the study, interpreted and analyzed the data, and drafted and revised the manuscript for intellectual content. LD, QL, FL, BW, YZ, QM, WL, JP, JX, SW, JY, HL, and JM provided the data of the validation cohort. JZ, WH, XC, HT, JY, and LZ provided the data of the primary cohort

and revised the manuscript for intellectual content. CL, MW, QD, JL, and CZ revised the manuscript for intellectual content. CQ designed and conceptualized the study, interpreted and analyzed the data, and revised the manuscript for intellectual content. LW and CQ conducted the statistical analyses of this manuscript. The corresponding author had full access to all the data in this study and had final responsibility for the decision to submit it for publication. All authors contributed to the article and approved the submitted version.

## Funding

This study was supported by the National Natural Science Foundation of China (Grant No. 82171341), the Shanghai Municipal Science and Technology Major Project (No. 2018SHZDZX01) and ZHANGJIANG LAB, and the National Key Research and Development Program of China (2016YFC0901504).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2022.947974/full#supplementary-material>

## References

1. Jarius S, Paul F, Weinshenker BG, Levy M, Kim HJ, Wildemann B. Neuromyelitis optica. *Nat Rev Dis Primers.* (2020) 6:85. doi: 10.1038/s41572-020-0214-9

2. Pittock SJ, Zekeridou A, Weinshenker BG. Hope for patients with neuromyelitis optica spectrum disorders - from mechanisms to trials. *Nat Rev Neurol.* (2021) 17:759–73. doi: 10.1038/s41582-021-00568-8

3. Poupard J, Giovannelli J, Deschamps R, Audoin B, Ciron J, Maillart E, et al. Evaluation of efficacy and tolerability of first-line therapies in NMOs. *Neurology*. (2020) 94:e1645–56. doi: 10.1212/WNL.00000000000009245
4. Traboulsee A, Greenberg BM, Bennett JL, Szczechowski L, Fox E, Shkrobot S, et al. Safety and efficacy of satralizumab monotherapy in neuromyelitis optica spectrum disorder: a randomised, double-blind, multicentre, placebo-controlled phase 3 trial. *Lancet Neurol*. (2020) 19:402–12. doi: 10.1016/S1474-4422(20)30078-8
5. Yamamura T, Kleiter I, Fujihara K, Palace J, Greenberg B, Zakrzewska-Pniewska B, et al. Trial of satralizumab in neuromyelitis optica spectrum disorder. *N Engl J Med*. (2019) 381:2114–24. doi: 10.1056/NEJMoa1901747
6. Pittock SJ, Berthele A, Fujihara K, Kim HJ, Levy M, Palace J, et al. Eculizumab in aquaporin-4-positive neuromyelitis optica spectrum disorder. *N Engl J Med*. (2019) 381:614–25. doi: 10.1056/NEJMoa1900866
7. Cree BAC, Bennett JL, Kim HJ, Weinshenker BG, Pittock SJ, Wingerchuk DM, et al. Inebilizumab for the treatment of neuromyelitis optica spectrum disorder (N-MOMentum): a double-blind, randomised placebo-controlled phase 2/3 trial. *Lancet*. (2019) 394:1352–63. doi: 10.1016/S0140-6736(19)31817-3
8. Cox DR. Regression models and life-tables (with Discussion). *J R Stat Soc Series B*. (1972) 34:187–220. doi: 10.1111/j.2517-6161.1972.tb00899.x
9. Wang L, Du L, Li Q, Li F, Wang B, Zhao Y, et al. Neuromyelitis optica spectrum disorder with anti-aquaporin-4 antibody: outcome prediction models. *Front Immunol*. (2022) 13:873576. doi: 10.3389/fimmu.2022.873576
10. Singal AG, Mukherjee A, Elmunzer BJ, Higgins PD, Lok AS, Zhu J, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. *Am J Gastroenterol*. (2013) 108:1723–30. doi: 10.1038/ajg.2013.332
11. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, et al. Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. *Circ Res*. (2017) 121:1092–101. doi: 10.1161/CIRCRESAHA.117.311312
12. Waljee AK, Higgins PD. Machine learning in medicine: a primer for physicians. *Am J Gastroenterol*. (2010) 105:1224–6. doi: 10.1038/ajg.2010.173
13. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. (2015) 521:436–44. doi: 10.1038/nature14539
14. Xia W, Hu B, Li H, Shi W, Tang Y, Yu Y, et al. Deep learning for automatic differential diagnosis of primary central nervous system lymphoma and glioblastoma: multi-parametric magnetic resonance imaging based convolutional neural network model. *J Magn Reson Imaging*. (2021) 54:880–7. doi: 10.1002/jmri.27592
15. González G, Ash SY, Vegas-Sánchez-Ferrero G, Onieva Onieva J, Rahaghi FN, Ross JC, et al. Disease staging and prognosis in smokers using deep learning in chest computed tomography. *Am J Respir Crit Care Med*. (2018) 197:193–203. doi: 10.1164/rccm.201705-0860OC
16. Kim DW, Lee S, Kwon S, Nam W, Cha IH, Kim HJ. Deep learning-based survival prediction of oral cancer patients. *Sci Rep*. (2019) 9:6994. doi: 10.1038/s41598-019-43372-7
17. Stellmann JB, Krumbholz M, Friede T, Gahlen A, Borisow N, Fischer K, et al. Immunotherapies in neuromyelitis optica spectrum disorder: efficacy and predictors of response. *J Neurol Neurosurg Psychiatry*. (2017) 88:639–47. doi: 10.1136/jnnp-2017-315603
18. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. (1996) 15:361–87. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4
19. Huang CC, Chan SY, Lee WC, Chiang CJ, Lu TP, Cheng SH. Development of a prediction model for breast cancer based on the national cancer registry in Taiwan. *Breast Cancer Res*. (2019) 21:92. doi: 10.1186/s13058-019-1172-6
20. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324
21. Ishwaran H, Kogalur U, Blackstone E, Lauer M. Random survival forests. *Annals of Applied Statistics*. (2008) 2:841–60. doi: 10.1214/08-AOAS169
22. Ehrlinger J. *ggRandomForests: Exploring Random Forest Survival*. Cambridge, MA (2016).
23. Kvamme H, Borgan Ø. Continuous and discrete-time survival prediction with neural networks. *Lifetime Data Anal*. (2021) 27:710–36. doi: 10.1007/s10985-021-09532-6
24. Lee C, Zame WR, Yoon J, Van Der Schaar M. DeepHit: a deep learning approach to survival analysis with competing risks. In: *32nd AAAI Conference on Artificial Intelligence*. New Orleans, LA: AAAI 2018 (2018), 2314–21. doi: 10.1609/aaai.v32i1.11842
25. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol*. (2018) 18:24. doi: 10.1186/s12874-018-0482-1
26. Mainali S, Darsie ME, Smetana KS. Machine learning in action: stroke diagnosis and outcome prediction. *Front Neurol*. (2021) 12:734345. doi: 10.3389/fneur.2021.734345
27. Kim H, Lee Y, Kim YH, Lim YM, Lee JS, Woo J, et al. Deep learning-based method to differentiate neuromyelitis optica spectrum disorder from multiple sclerosis. *Front Neurol*. (2020) 11:599042. doi: 10.3389/fneur.2020.599042
28. Palace J, Lin DY, Zeng D, Majed M, Elson L, Hamid S, et al. Outcome prediction models in AQP4-IgG positive neuromyelitis optica spectrum disorders. *Brain*. (2019) 142:1310–23. doi: 10.1093/brain/awz054
29. Kunchok A, Malpas C, Nytrova P, Havrdova EK, Alroughani R, Terzi M, et al. Clinical and therapeutic predictors of disease outcomes in AQP4-IgG+ neuromyelitis optica spectrum disorder. *Mult Scler Relat Disord*. (2020) 38:101868. doi: 10.1016/j.msard.2019.101868
30. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods*. (2009) 14:323–48. doi: 10.1037/a0016973