

High Dimensional Survival Analysis Model for Predicting Diagnosis of alzheimer's disease over Specific Period of time

First Author^{1,2*}, Second Author^{2,3†} and Third Author^{1,2†}

^{1*}Department, Organization, Street, City, 100190, State, Country.

²Department, Organization, Street, City, 10587, State, Country.

³Department, Organization, Street, City, 610101, State, Country.

*Corresponding author(s). E-mail(s): iauthor@gmail.com;

Contributing authors: iiauthor@gmail.com; iiiauthor@gmail.com;

[†]These authors contributed equally to this work.

Abstract

Survival analysis is a statistical tool for predicting the period until an event such as death or disease diagnosis. It requires the availability of censored data demonstrating that the event of interest did not occur during the study period. The model faces issues of overfitting if it has more features than data. However, not all elements are necessary for addressing the problem, and incorporating non-essential aspects might occasionally degrade learning performance. As a result, constructing an accurate survival model using electronic health records is difficult. With this rationale, a hybrid approach for high-dimensional survival analysis is developed. First, labeled (Alzheimer's Disease) and unlabeled (Normal Cognitive) instances are used to offer better representation with lower dimensions from clinical features, and then we actively train the survival model by labeling the censored data. The effectiveness of this strategy was evaluated using a c-index study. The results show that our method outperforms baseline models by a wide margin.

Keywords: survival analysis,time-series,AD,feature

1 Introduction

Dementia is a clinical illness marked by progressive declines in various cognitive areas (e.g., memory and executive function) that interfere with daily social and professional functioning [? ?]. Alzheimer's disease is the most frequent kind of dementia, followed by vascular dementia, accounting for 50–70% and 15–20% of all dementia cases, respectively. As it progresses, symptoms such as disorientation, confusion, and behavioral changes become increasingly severe. Speaking, swallowing, and walking grows difficult with time [?]. Alzheimer's disease is the sixth most common cause of mortality in the United States, accounting for 5% of deaths. It is the fifth most significant cause of death for people aged 65 and up [?]. In the absence of a medical breakthrough, the number of individuals living with dementia globally is approximately 50 million, with that number expected to triple by 2050.

Alzheimer's disease is likely to start 20 years or more before symptoms appear, with undetectable changes in the brain. Individuals only start to notice signs like memory loss and language problems after years of brain changes [? ?]. As a result, identifying patients at high risk of developing Alzheimer's disease is critical [? ?]. At the same time, early detection is critical for developing a treatment approach that would reduce the progression. It is when the sickness progresses from one symptom to the next. At the same time, current research is mainly focused on forecasting whether it will transition to a different stage.

Data from dementia studies is usually high-dimensional and censored but heterogeneous, with data originating from various sources with varied statistical features and missing data. A recent study has demonstrated that several sources of clinical data can provide complementary information concerning dementia and that combining many sources of data improves cognitive decline prediction over using a single source [?]. It poses analytical difficulties, such as the significant danger of overfitting the model to the data, restricting the model's capacity to generalize to new data, and the high variance of models fitted to this data. As a result, methods to tackle the obstacles posed by high-dimensional clinical data are urgently needed.

Feature selection techniques have become critical components of the learning process in order to deal with the problem of excessive data dimensionality [? ?]. As a result, selecting the right features can help the inductive learner increase their learning speed, generalization capacity, and induced model simplicity. Feature selection and feature extraction are the two most used dimensionality reduction strategies, each with its own set of advantages [?]. Feature extraction algorithms combine the original features to reduce dimensionality. In contrast, Feature selection reduces dimensionality by deleting irrelevant and redundant features. As a result, selecting the right features can help the inductive learner increase their learning speed, generalization capacity, and induced model simplicity. When modeling high-dimensional, heterogeneous clinical data, machine learning models that forecast the period before a patient develops dementia are crucial tools. It helps identify dementia risks and can provide more accurate results than standard statistical methods.

Survival analysis is a statistical tool for predicting the time until an event such as death, disease diagnosis, or mechanical component failure. The availability of censored data, showing that the event of interest did not occur within the research period, is an essential component of survival analysis. The availability of censored data necessitates the application of specialized methods. The Cox, a proportional hazards model, has long been the most popular method for analyzing censored data, although it was created for small data sets and did not scale well to large dimensions. Machine learning techniques with high-dimensional data have worked with censored data, allowing machine learning to provide more flexible options.

This paper aims to propose a suitable framework for high-dimensional and heterogeneous clinical data. It is responsible for cognitive aging and dementia by predicting its survival time. It is organized as follows: Section 2 discusses the related work presented in the literature for identifying AD. Section 3 describes the scientific approaches, methods, and data. Section 3.5 explores the experimental to illustrate the results. At the same time, the discussion of the experimental is shown in section 4. Finally, Section 5 concludes the findings and results of the paper.

2 Related Work

Yang et al. [?] devised a functional linear regression model and used it to estimate conversion time and show that CC atrophy increases AD development using ADNI data. It shows that the suggested model can appropriately restore the failure time in right censoring and a functional covariate using simulation studies. Finally, they discovered that CC's atrophy in the back enhances Alzheimer's disease progression.

Casanova et al. [?] use MRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) to investigate the feasibility of estimating an anatomical index. This index was given the name AD Pattern Similarity (AD-PS) scores. It can be used as an Alzheimer's disease (AD) risk factor in the Women's Health Initiative Magnetic Resonance Imaging Study (WHIMS-MRI). Then ADNI data is used to estimate a novel AD risk factor using a high-dimensional machine learning approach. The GM AD-PS scores exhibited substantial cross-sectional and longitudinal relationships with age, cognitive function, cognitive status, and WM SVID volume, indicating that the GM AD-PS score is still being validated.

Most Alzheimer's disease (AD) research has relied only on medical imaging. Mart-Juan et al. [?] conducted a study that focused on longitudinal imaging data. It concentrated on papers published between 2007 and 2019. Long short-term memory (LSTM) is used by Hong et al. [?] to predict the progression of Alzheimer's disease. It predicts the disease's future condition rather than the state of a current diagnosis.

Janghel [?], on the other hand, develops and examines several approaches for diagnosing and predicting Alzheimer's disease using only MRI images. It

only uses one model: a convolutional neural network (CNN). It employs four different CNN architectures at the same time. A support vector machine is used to find the best subsets of features for binary classification (SVM). Cai et al. [?] offers an embedded feature selection approach based on the least-squares loss function and within-class scatter for determining the optimal feature subset.

Bi et al. [?] also talked about deep learning technologies. It concentrated on the automatic detection of Alzheimer's disease using MRI images. It follows a two-step process: 1-for feature extraction, use the unsupervised CNN. 2-reaches a final diagnosis using an unsupervised predictor. Based on volumetric features from sMRI data, Bashe et al. [?] proposed an aggregated strategy using Hough-CNN, CNN, and DNN models to diagnose Alzheimer's disease. The proposed DNN model utilizes the volumetric features retrieved by the DVE-CNN model to classify the AD and NC classes.

Grassi et al. [?], Liu et al. [?] is the only study that uses more realistic and cheap data for diagnosis. To anticipate three-year conversion [?] uses a weighted rank average gathered by multiple supervised machine learning approaches. Liu et al. [?] propose a new approach for identifying AD using spectrogram characteristics derived from voice recordings. A small number of criteria are used to make predictions (socio-demographic, clinical, and neuropsychological scores). The significant benefit is the use of algorithmic decision-making tools.

Use the appropriate correlation analysis approach at the end to uncover relationships between brain areas and genes. A cluster evolutionary random forest was used to propose [?] (CERF). To improve the random forest's generalization performance, it introduces the idea of clustering evolution. Farouk and Rady [?] looked at unsupervised clustering algorithms for Alzheimer's disease early detection. Although classification algorithms are used to identify medical illnesses, the lack or inaccuracy of labeled data may be a problem. Using Voxel-Based Morphometry (VBM) characteristics extracted from MRI images, this study compares the k-means and k-medoids.

The following studies serve as the foundation for our study. In ascending order, they are listed. The first, [?], explains how MRI data can improve the accuracy of diagnoses for the Mini-Mental State Examination (MMSE) and logical memory (LM) tests. It accesses model correctness via Multilayer Preceptor. The second, [?], shows how clinically translatable strategies for conversion can be predicted. It also detects high-risk people who are converted. It continues to work three years after the initial assessment. Then, Haaksma et al. [?] address the link between Alzheimer's disease and its predictors. It included some Alzheimer's disease cases that have had at least one examination following diagnosis. To determine whether there are any latent classes of MiniMental State Examination (MMSE) and Clinical Dementia Rating sum of boxes (CDRsb) routes across time. It employs growth mixture models with parallel processes. To find baseline predictors of class membership, researchers utilised bias-corrected multinomial logistic regression. A multimodal data [?] classifier that employs a hybrid deep neural network classifier. It is based on a

set of MRI pictures as well as EEG inputs. The goal is to improve the learning process by incorporating the weight component of DNN into CNN. Then it explains how the accuracy of hybrid classifiers is determined.

3 Materials & Methods

3.1 Data preparation

Principal Investigator Michael W. Weiner, MD, founded the ADNI in 2003 as public-private cooperation. The primary purpose of ADNI was to see if various variables could be integrated to track moderate cognitive impairment (MCI) and early Alzheimer's disease progression (AD). Magnetic resonance imaging (MRI), positron emission tomography (PET), various biological indicators, and clinical and neuropsychological evaluation are among the aspects. Also, it includes various information, such as demographics, APOE gene status, neuropsychological test scores, medical history, family history, medical examination, blood test results, and adverse events. The characteristics of study is summarised in Table 1.

Table 1 Study characteristics

Study Design	Longitudinal, observational clinical trial
Sample size(n)	12612
Number of features (p)	90
Censoring rate (%)	
Follow-up period	84 month
Age at baseline	55–90 years

3.2 Subjects

Our prediction model was trained and tested on data extracted from the ADNI with 18 month longitudinal trajectories of 900 cases on each class (NL, AD). It covers of total 1800 records. It covers 24 different neurological test with corresponding MRI. Each patient profile consisted of 24 test and 7 images files. Patient trajectories described the time evolution of all variables in 3-month intervals. Detailed data processing steps are described in next subsection. The following are the subject criteria that were employed in this study: 1- Age ranges from 55 to 90; 2-Education levels range from primary to graduate; 3- All colours and ethnicities included. The following are the different sorts of data that were used: 1-Neurological test (neuropsychologist), 2-Baseline: initial tests and diagnoses for patients, and 3-Brain image technology (MRI only)

3.3 Data Preprocessing

Data pre-processing get the data ready for the machine learning algorithms. This study assumes that data are missing at random and that censoring is non-informative, i.e., it is unaffected by a dementia diagnosis. Missing values

are common in clinical datasets, and imputation is the process of substituting appropriate values for the ones that are missing. Multiple imputations, which preserve the data's relationships and the uncertainty in those relationships, commonly impute missing data.

It includes participants with cognitive typical (NL), mid-cognitive impairment (MCI), and Alzheimer's Disease (AD) who were recruited from over 50 different US and Canadian centers with six-month follow-up examinations. Although there are nine classes, only two classes (AD and NL) demonstrate the concept of AD survival time prediction. The data ranged from Imaging (MRI), time, neurological test data collected using Assessment data.

The prediction matrix built on the training set was used to perform imputation within the cross-validation cycle, first on the training set and then on the test set. After data preprocessing, the ADNI dataset contained 12 features, one categorical, three ordinal and the rest numeric (image is converted to numeric). All continuous characteristics applied the normalization process. Any features removed that had more than 60% of their values missing. Table 2 shows the data result after preprocessing.

Table 2 Data summary after Preprocessing step

Name	Type	Mean
DX bl	categorical	–
CDRSB	ordinal	3.578
ADAS11	ordinal	16.64
ADAS13	ordinal	25.474
RAVLT immediate	numeric	25.192
RAVLT learning	numeric	2.43
RAVLT forgetting	numeric	4.451
FAQ	numeric	10.018
Hippocampus	numeric	5936.106
Years bl	numeric	1.043
Month	numeric	12.35
Month bl	numeric	12.49

3.4 Model Selection

The number of attributes in a dataset is called dimensionality in statistics. Healthcare data, for example, is known for having a large number of variables (e.g., blood pressure, weight, cholesterol level). It is represented in a spreadsheet in an ideal world, with one column for each dimension. The main issue is the performance in practice because many variables are interconnected (like weight and blood pressure). The term "high dimensional" refers to a situation in which the number of dimensions is so large that computations become highly complicated. The number of characteristics in high-dimensional data can outnumber the number of observations.

In general, there are ten machine learning algorithms and eight feature selection methods capable of handling high-dimensional and heterogeneous data. The machine learning algorithms is divided into three categories [?]:

- Penalised Cox Regression: LASSO, ElasticNet and Ridge regression.
- Boosted Cox Regression: Cox model with likelihood based boosting (Cox-Boost), Cox model with gradient boosting (GLMBoost) and Extreme Gradient Boosting (XGBoost) with tree-based and linear model-based boosting.
- Random Forests: random survival forest, maximally selected rank statistics random survival forest.

The feature selection methods divided into categories [?]:

- Filter Methods: univariate Cox score (Univariate), random forest variable importance (RF Var Imp), random forest minimal depth (RF Min Depth), random forest variable hunting (RF Var Hunt), maximally selected rank statistics random forest (RF Max Stat), minimum redundancy maximum relevance (mRMR).
- Wrapper Methods: sequential forward selection (SFS), sequential forward floating selection (SFFS).

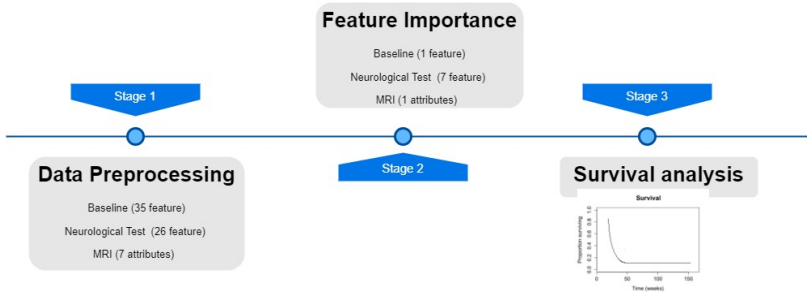


Fig. 1 Proposed Model pipeline

This paper proposes a general survival ensembles framework based on random forests using a split criterion. Our approach uses RF Var Imp and Cox-Boost. Figure 1 illustrates the proposed model pipeline. The phases depends on each other. The following summarizes the steps of the different parties within the proposed framework pipeline.

1. Data Preprocessing is where data quality improves to extract relevant insights from it. In simple terms, it is a data mining approach used in Machine Learning that turns raw data into a legible and intelligible format. (For more information about these steps, see the previous subsection)
2. Feature Importance refers to strategies that give input features a score based on predicting a target variable. The original data contains 68 attributes, while after this step, there are nine attributes.
3. Survival analysis entails modeling time to event data; in the survival analysis literature, an "event" is as death or failure in this sense. The event in this work is an AD diagnosis.

It adapted to censored data and then derived variable selection strategies for high-dimensional survival data. The population level was used to create the model. A change in AD disease diagnosis at any moment during the study period was the event of interest.

This paper presents a methodological framework for analyzing AD and survival data using RF Cox regression methods that have been cross-validated. The procedure begins with raw data analysis and then takes the reader through the evaluation of the conclusions using a cross-validated penalty approach. As shown in Figure 2, the general steps of our approach are the following: (i) The model uses three different data types. One is an MRI image, while the others are neurological test results and diagnosis of the first visit. It contains 68 features (26 neurological tests, 7 MRIs, and 35 baselines); Where each one represents one patient record. The first stage in the proposed model got the patients' data as input values. This stage deals with the missing values. (ii) Feature importance, where the best feature is selected. Including all this type of data costs a lot of money and processing power. As a result, the proposed model chooses the best features to increase performance. This output is the attribute that best describes patients' survival rate. (iii) Using CoxBoost models to analyze Alzheimer's disease; (iv) Assessing survival models to estimate the prognosis of Alzheimer's patients. The methodological framework given here provides a novel scientific framework for the study and interpretation of regression methods that may easily be extended to include different Cox algorithms.

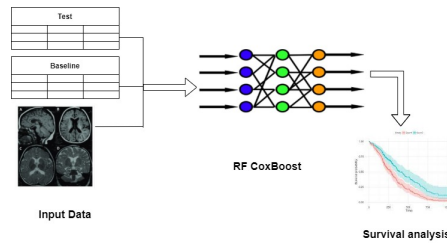


Fig. 2 Proposed Model Framework

Many machine learning algorithms contain one or more hyper-parameters that must be chosen to optimize model performance. A 10-fold nested cross-validation loop was used to tune these hyper-parameters automatically. In the inner loop, values for the hyperparameters were chosen using a random search with 25 iterations, and model performance was assessed in the outer loop. All model selection processes were then performed for each training/test data combination in this manner.

3.5 Model Evaluation

The concordance index was used as a criterion (C-Index) to evaluate the models. It counts how many pairs of observations have a higher survival probability

than the model predicts. When random sampling, such as cross-validation, is used, random partitioning might affect model performance. As a result, the model performance findings are subjected to statistical significance tests.

It considers a censored time-to-event survival model, such as the Cox regression model. A model's performance in terms of discrimination and calibration is assessed. Discrimination refers to a model's capacity to appropriately differentiate between two types of outcomes. Individuals with events had higher predicted probabilities than subjects who did not have events when using a model with good discrimination ability[?].

The most used survival model evaluation metrics is the Concordance index (CI) [? ? ?] which is the generalization of the ROC curve in the complete data in the survival data [? ?]. CI can be divided into two categories according to whether considering the data tied.

In practical applications, deaths may be observed in both samples ($\delta_i=1$ and $\delta_j=1$) in the sample pair, and the observed survival times are the same $T_i = X_i = X_j = T_j$. In this case, when the survival probabilities or survival times predicted by the model are also equal $Y_i = Y_j$, the model has a perfect predictive ability. However, without considering ties between survival times, this kind of sample pair cannot improve CI for the original CI definition.

To deal with this problem, Ishwaran introduced an improved CI calculation method for the tied survival data [?]. According to their definition, if the prediction result is survival time or survival probability, the smaller the prediction result, the worse it is; if the prediction result is the hazard function, the larger the value, the worse the prediction result. Detailed information on how to calculate CI are shown below:

- First define the comparable sample pair:

$$np_{ij}(X_i, \delta_i, X_j, \delta_j) = \max(I(X_i \geq X_j)\delta_j, I(X_i \leq X_j)\delta_i) \quad (1)$$

- In the second step, calculate the Complete Concordance (CC):

$$CC = \sum_{ij} I(\text{sign}(Y_i, Y_j) = \text{csign}(X_i, X_j) \mid np_{ij}) \quad (2)$$

where

$$\begin{aligned} \text{sign}(Y_i, Y_j) &= I(Y_i \geq Y_j) - I(Y_i \leq Y_j) \\ \text{csign}(X_i, \delta_i, X_j, \delta_j) &= I(X_i \geq X_j)\delta_j - I(X_i \leq X_j)\delta_i \end{aligned}$$

- Then, derive the Partial Concordance (PC):

$$\begin{aligned} PC &= \sum_{ij} I(Y_i = Y_j \mid np_{ij} = 1, X_i \neq X_j) \\ &\quad + I(Y_i \neq Y_j \mid np_{ij} = 1, X_i = X_j, \delta_i = \delta_j = 1) \\ &\quad + I(Y_i \geq Y_j \mid np_{ij} = 1, X_i = X_j, \delta_i = 1, \delta_j = 0) \end{aligned} \quad (3)$$

- Finally, CI can be calculated as:

$$CI = \frac{\text{Concordance}}{\text{Permissible}} = \frac{CC + 0.5 * PC}{\sum_{i,j} np_{ij}(X_i, \delta_i, X_j, \delta_j)} \quad (4)$$

where, δ_j is the survival status of sample i (0 means censoring, 1 means event), Y_i and X_i represents the predicted survival time and the observed survival time, respectively.

Perfect discrimination would result in the model producing two non-overlapping sets of projected probabilities: positive outcomes and adverse events. The degree to which the anticipated probability corresponds numerically with the actual events is calibration [?]. The model is well-calibrated when the expected and observed values agree for any plausible grouping of the observations, ordered by increasing predicted values. Calibration measures are statistics that divide a data set into groups and compare the average expected probability to the prevalence of the result in each category.

The null hypothesis uses paired t-test to test that the mean of two sets of values is the same. Because the numerous training and test sets may overlap, the t-independence test's assumption is violated when the two sets of values in question are the performance results of two models tested using random sampling, such as repeated k-fold cross-validation.

The experiments happened using the following conditions. The R package mlr (Machine Learning in R) was used as a framework to carry out benchmarking, and all code for the experiments was written in R. Also, R package survival and survminer used to illustrates survival analysis for AD patients. All resampling was performed using five repeats of stratified 10-fold cross-validation. The data is divided into 70% training data and 30% testing data.

4 Results

A Cox model is a widely used statistical method for assessing the relationship between a patient's survival and several explanatory variables. A Cox model predicts the treatment effect on survival after correcting for other explanatory variables [? ?]. In this paper, survival discussed the patient being diagnosed with AD. These terms imply that the estimate ignores the shape of the survival function. On the other hand, a histogram is an empirical evaluation of a variable's distribution, whereas a standard curve presupposes that the distribution has a fixed shape. Table 3 shows the estimates of performance metrics for the fittings of the model. It computes the predicted survivor function for AD patient using Cox proportional hazards model. This table illustrated the following:

- time: follows patients through 48 months.
- n.risk: the number of patients at risk of AD diagnosis each time. It started with 1800 patients, and then it decreased to its minimum at the end of time

interval (48 months) to 67. This means that there are 67 patients at risk of developing AD.

- **n.event**: there are 239 events that occurred at the beginning. Then this number start decreasing until it reaches 27 at the end of time.
- **survival**: estimate of survival probability. The whole number of patients had 86% of developing disease. This number start decreasing to 16% after 48 month.
- **std.err**: standard error of survival. It got a low percentage of survival error. For 1800 patients there are an error of 0.8% of error. While, for 67 patient it became 1.8%, it is a low percentage of error for predicting patients progression.
- **lower and upper 95% CI**: lower end of confidence interval. Approximately 95% of the intervals produced would capture the actual population mean if the sampling process was repeated multiple times. As a result, as the sample size grows, the range of interval values narrows, determining the mean with greater precision than with a smaller sample. The probability of the population mean value being between -0.852 and 0.883 standard deviations (z-scores) from the sample mean 95% for a 0-time interval. While for a 48-month interval is between -0.126 and 0.200. As a result, 5% of the population's mean risk of falling beyond the upper and lower confidence intervals.

Table 3 Survival Fit Model Summary

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
0	1800	239	0.867	0.0080	0.852	0.883
6	1314	212	0.727	0.0111	0.706	0.749
12	910	167	0.594	0.0130	0.569	0.620
18	576	123	0.467	0.0144	0.440	0.496
24	409	79	0.377	0.0148	0.349	0.407
36	181	53	0.266	0.0165	0.236	0.301
48	67	27	0.159	0.0188	0.126	0.200

Cox regression, Cox model, and Cox proportional hazards regression are all terminology used to describe a semi-parametric method developed by Cox [?]. The model is semi-parametric because it makes no assumptions about the event time distribution. However, it does assume that the hazard function depends on a set of parameters. These parameters characterize the relationship between the hazard and a set of predictors. The Cox model is expressed by the hazard function denoted by $h(t)$. In a nutshell, the hazard function represents the danger of dying at time t . It is possible to estimate it as follows:

$$h(t) = h_0(t) * \exp(b_1x_1 + b_2x_2 + \dots + b_px_p) \quad (5)$$

where,

- t represents the survival time

- $h(t)$ is the hazard function determined by a set of p covariates $(x_1, x_2, \dots, x_p)$
- the coefficients $(b_1, b_2, \dots, b_p)$ measure the impact (i.e., the effect size) of covariates.
- the term h_0 is called the baseline hazard. It corresponds to the value of the hazard if all the x_i are equal to zero (the quantity $\exp(0)$ equals 1). The t in $h(t)$ reminds us that the hazard may vary over time.

Table 4 use Cox regression to numerically assess the magnitude of different disparities by estimating the hazard ratios comparing these groups. The groups used depends on the year. The Chi-Square test of independence performs using the following hypotheses: H0: Years less than 3 and gender preference are independent. H1: Years greater than 2 and gender preference are not independent. The function returns a list of components, including:

- n : the number of patients in each group. 1482 patient for cases distributed less than 2 years, and 318 patients in more than two years
- obs : for the first group there are 770 events; while the other group there are 130 events.
- exp : 539 and 361 is the weight expect number of patient diagnosis in first group and second group respectively.
- $(O - E)^2/E$ for each group depends on each test. For first hypothesis there are 99.2; while the other test got 148.0.
- V is the theoretical variance of O-E for the first group. If O-E is approximately normally distributed, as it will be in large samples, then $(O - E)^2/V$ will be approximately chi-squared distributed on 1 DF

Here, the number of patients in each group equals the corresponding observed number of events since there is no censoring. The value of chi-square statistics is 415 on 1 degree of freedom, and the p-value is 2e-16, which is not statistically significant. It is used to compare observed results with expected results. This test aims to determine whether the difference between observed data and expected data is due to chance or if it is due to a relationship between the variables studied. The log-rank test for difference in survival gives a p-value of $p = 2e-16$, indicating that grouping the year differs significantly in survival. If the P-Value is less than 0.05, the null hypothesis is rejected in the test with 95% confidence or 5% significance. In general, we reject the null hypothesis when the P-value is less than the level of significance of the test.

Table 4 Survival difference Model Summary

	N	Observed	Expected	$(O - E)^2/E$	$(O - E)^2/V$
0	1482	770	539	99.2	415
1	318	130	361	148.0	415

5 Discussion

The time for events to occur, referred to as survival time, is studied and modeled in survival analysis. The Cox proportional-hazards regression model is the most frequent method for examining the relationship between survival time and predictor factors. Table 5 shows the effect of each features on the survival analysis. It consist of the following

- `coef` : measure the impact of covariates(Log Hazard Ratio)
- `exp(coef)`: Hazard Ratio, give the effect size of covariates.
- `se(coef)`: Standard Error
- `z`: It gives the Wald statistic value, where it corresponds to the ratio of each regression coefficient to its standard error ($z = \text{coef}/\text{se}(\text{coef})$).
- `Pr(¿z)`: Probability of Wald statistic value. It views the significance of each feature. DX_ bl for CN, DX_ bl for LMCI, CDRSB are highly significant values (There values affect the survival results). Then comes FAQ with lower significance.
- The hazard ratios' confidence intervals: The hazard ratio (`exp(coef)`) has upper and lower 95 percent confidence intervals in the summary output.

Table 5 Cox proportional-hazards regression model Summary

	coef	exp(coef)	se(coef)	z	Pr(¿z)	lower 0.95	upp
DX blCN	6.605e+00	7.390e+02	1.101e+00	5.998	2.00e-09	85.3637	639
DX blEMCI	-8.691e+00	1.681e-04	1.332e+03	-0.007	0.995	0.0000	Inf
DX blLMCI	6.155e+00	4.712e+02	1.006e+00	6.117	9.54e-10	65.5666	338
CDRSB	-3.263e-01	7.216e-01	5.222e-02	-6.249	4.14e-10	0.6514	0.79
ADAS11	-5.029e-03	9.950e-01	2.323e-02	-0.217	0.829	0.9507	1.04
ADAS13	-6.066e-03	9.940e-01	1.820e-02	-0.333	0.739	0.9591	1.03
RAVLT immediate	3.398e-03	1.003e+00	6.243e-03	0.544	0.586	0.9912	1.01
RAVLT learning	-2.788e-03	9.972e-01	2.105e-02	-0.132	0.895	0.9569	1.03
RAVLT forgetting	2.370e-02	1.024e+00	1.903e-02	1.246	0.213	0.9865	1.06
FAQ	-2.420e-02	9.761e-01	1.207e-02	-2.005	0.045	0.9533	0.99
Hippocampus	-3.840e-05	1.000e+00	3.995e-05	-0.961	0.336	0.9999	1.00
Years bl	5.213e+03	Inf	1.709e+04	0.305	0.760	0.0000	Inf
Month bl	-4.359e+02	4.716e-190	1.427e+03	-0.305	0.760	0.0000	Inf

The likelihood-ratio test, the Wald test, and score log-rank statistics are among the three possible tests for the overall significance of the model. Asymptotically, these three techniques are equivalent. They will get comparable outcomes if N is large enough. They may differ slightly for small N. The likelihood ratio test performs better with small sample sizes used commonly. As shown in table 5 it got the following results:

- Concordance= 0.998 (se = 0.001)
- Likelihood ratio test= 2727 on 13 df, $p=¿2e-16$
- Wald test = 583.5 on 13 df, $p=¿2e-16$
- Score (logrank) test = 1419 on 13 df, $p=¿2e-16$

When, It comes to evaluating the performance of the model. It means that the model got better performance when it combined multiple features. It means that combining more than one feature for a survival model will get better performance.

The most frequently used evaluation metric of survival models is the concordance index (c index, c statistic). The concordance index or c-index is a metric to evaluate the predictions made by model. It is defined as the proportion of concordant pairs divided by the total number of possible evaluation pairs [?].

$$C - index = \frac{N.ConcordantPair}{N.ConcordantPair + N.DiscordantPair} \quad (6)$$

Where

- Concordant pairs: both interviewers rank both applicants in the same order — that is, they both move in the same direction. While they aren't the same rank (i.e. both 1st or both 2nd), each pair is ordered equally higher or equally lower. Interviewer 1 ranked F as 6th and G as 7th, while interviewer 2 ranked F as 5th and G as 8th. F and G are concordant because F was consistently ranked higher than G.
- Discordant pairs: Candidates E and F are discordant because the interviewers ranked in opposite directions (one said E had a higher rank than F, while the other said F ranked higher than G).

What is a good C index? The area under the Receiver Operating Characteristic (ROC) curve is equal to the image result for the c index, which runs from 0.5 to 1. A value of 0.5 indicates that the model is no better than random chance in predicting a result. A model with a value greater than 0.7 is considered good. A value of less than 0.5 suggests a poor model. Table 6 shows c-index value for both proposed model and Random Forests for Survival. Our proposed model illustrates that the model has a value near 1 (0.953) which mean a good model with great performance. It outperform other survival model.

Table 6 Model Performance using c-index

Proposed Model	Random Forests for Survival
0.953	0.902

Our research is based on the assumption that each time-dependent variable in a patient's clinical record is stochastic. It is sampled from a range of values rather than taking on a single deterministic value. Figure 3 shows the Kaplan-Meier estimates of the survival curves following AD diagnosis.

6 Conclusion

This paper introduces a deep learning model for Survival analysis of Alzheimer's disease. Because the condition is essentially progressive, the model

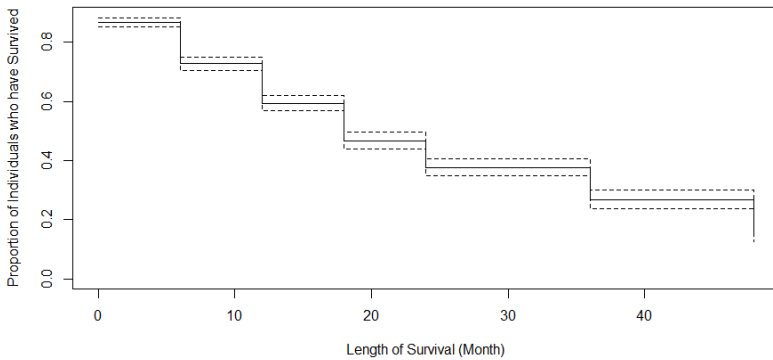


Fig. 3 Plot of Survival Curve for AD Patients

considers the timing data gathered from the cases. In contrast to previous methodologies, our model can predict the patient survival condition rather than classify the state of a current diagnosis. Experiments have shown that our model outperforms the traditional Cox Model. Improving the model's performance will require further research in future work. Personal data could improve the performance and efficiency of AD prediction at an earlier stage. Furthermore, testing the proposed methodology using actual data is required in further work.

Declarations

Some journals require declarations to be submitted in a standardised format. Please check the Instructions for Authors of the journal to which you are submitting to see if you need to complete this section. If yes, your manuscript must contain the following sections under the heading 'Declarations':

- Funding
- Conflict of interest/Competing interests (check journal-specific guidelines for which heading to use)
- Ethics approval
- Consent to participate
- Consent for publication
- Availability of data and materials
- Code availability
- Authors' contributions

If any of the sections are not relevant to your manuscript, please include the heading and write 'Not applicable' for that section.

Editorial Policies for:

Springer journals and proceedings:

<https://www.springer.com/gp/editorial-policies>

Nature Portfolio journals:

<https://www.nature.com/nature-research/editorial-policies>

Scientific Reports:

<https://www.nature.com/srep/journal-policies/editorial-policies>

BMC journals:

<https://www.biomedcentral.com/getpublished/editorial-policies>