**Supplementary Materials:**

**S1. Data cleaning procedures:**

We do the variable pre-processing by the below steps.

1. Free text fields were removed.
2. Administration variables were removed, e.g., patient ID, Form version, etc.
3. Cross-sectional variables that can make information leakage were removed, e.g., variables denoting if the patient finally progresses to clinical AD dementia or other dementia.
4. Codes in variables that denote missing values (e.g., -4, 8888, 9999, etc.) were converted to NAs
5. Variables with over 50% NAs were removed from the analysis.
6. In categorical variables, levels that only have limited observations were combined, e.g., American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Asian, Other (specify) are combined into others against white, and black.
7. For scale variables which also have indicator variables, we multiplied them together. For example, the delusion severity was calculated by (delusion severity level) \* (delusion indicator)
8. For those patients who are not even able to do the cognitive test due to cognitive problems, we denoted it as the worst level or score.
9. For continuous variables that denote cognition decline date or age, we convert them into categories to incorporate those who do not decline as zeros.
10. No dumpy variable conversion and normalization were used since our prediction method is the random forest which can easily handle categorical variables and is invariant for normalization.

**S.2 Subgroup Analysis for biomarker verified cohort**

**Table S1 Model performance of subgroup analysis**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type | Cohort | Cut-Month | AUC  at Cut-Month | Accuracy  at Cut-Month | C-index |
| Internal  validation | NACC-v1v2 | 48 | 93.55%  SE = 4.98% | 91.48%  SE = 4.26% | 85.76%  SE = 6.19% |
| External  validation | NACC-v3 | 48 | 89.85%  SE = 2.39% | 83.27%  SE = 2.40% | 85.76%  SE = 1.22% |

SE: standard error; the SE was obtained from ten-fold cross-validation or bootstrap sampling.

**S.3 Sensitivity Analysis for missing code 96**

**Table S2 Model performance of sensitivity analysis**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type | Cohort | Cut-Month | AUC  at Cut-Month | Accuracy  at Cut-Month | C-index |
| Internal  validation | NACC-v1v2 | 48 | 90.08%  SE = 1.18% | 89.88%  SE = 0.33% | 84.96%  SE = 1.27% |
| External  validation | NACC-v3 | 48 | 92.39%  SE = 1.12% | 88.52%  SE = 1.02% | 90.72%  SE = 0.72% |

# SE: standard error; the SE was obtained from ten-fold cross-validation or bootstrap sampling.

**S.4 Names of 317 Candidate variables in NACC**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| NACCREAS | NACCREFR | SEX | HISPANIC | HISPOR |
| RACE | RACESEC | RACETER | PRIMLANG | EDUC |
| MARISTAT | NACCLIVS | INDEPEND | RESIDENC | HANDED |
| INSEX | INHISP | INHISPOR | NACCNINR | INRACE |
| INRASEC | INRATER | INEDUC | INRELTO | INLIVWTH |
| INVISITS | INCALLS | INRELY | NACCFAM | NACCMOM |
| NACCDAD | NACCFADM | NACCFFTD | ANYMEDS | TOBAC30 |
| TOBAC100 | SMOKYRS | PACKSPER | QUITSMOK | CVHATT |
| CVAFIB | CVANGIO | CVBYPASS | CVPACE | CVCHF |
| CVOTHR | CBSTROKE | NACCSTYR | CBTIA | NACCTIYR |
| PD | PDYR | PDOTHR | PDOTHRYR | SEIZURES |
| NACCTBI | TRAUMBRF | TRAUMEXT | TRAUMCHR | NCOTHR |
| DIABETES | HYPERTEN | HYPERCHO | B12DEF | THYROID |
| INCONTU | INCONTF | ALCOHOL | ABUSOTHR | DEP2YRS |
| DEPOTHR | PSYCDIS | HEIGHT | WEIGHT | BPSYS |
| BPDIAS | HRATE | VISION | VISCORR | VISWCORR |
| HEARING | HEARAID | HEARWAID | ABRUPT | STEPWISE |
| SOMATIC | EMOT | HXHYPER | HXSTROKE | FOCLSYM |
| FOCLSIGN | HACHIN | STROKCOG | PDNORMAL | SPEECH |
| FACEXP | TRESTFAC | TRESTRHD | TRESTLHD | TRESTRFT |
| TRESTLFT | TRACTRHD | TRACTLHD | RIGDNECK | RIGDUPRT |
| RIGDUPLF | RIGDLORT | RIGDLOLF | TAPSRT | TAPSLF |
| HANDMOVR | HANDMOVL | HANDALTR | HANDALTL | LEGRT |
| LEGLF | ARISING | POSTURE | GAIT | POSSTAB |
| BRADYKIN | MEMORY | ORIENT | JUDGMENT | COMMUN |
| HOMEHOBB | PERSCARE | CDRSUM | CDRGLOB | COMPORT |
| CDRLANG | NPIQINF | DELSEV | HALLSEV | AGITSEV |
| DEPDSEV | ANXSEV | ELATSEV | APASEV | DISNSEV |
| IRRSEV | MOTSEV | NITESEV | APPSEV | NOGDS |
| SATIS | DROPACT | EMPTY | BORED | SPIRITS |
| AFRAID | HAPPY | HELPLESS | STAYHOME | MEMPROB |
| WONDRFUL | WRTHLESS | ENERGY | HOPELESS | BETTER |
| NACCGDS | BILLS | TAXES | SHOPPING | GAMES |
| STOVE | MEALPREP | EVENTS | PAYATTN | REMDATES |
| TRAVEL | NACCNREX | FOCLDEF | GAITDIS | EYEMOVE |
| DECSUB | DECIN | DECCLIN | COGMEM | COGJUDG |
| COGLANG | COGVIS | COGATTN | COGOTHR | NACCCOGF |
| COGMODE | DECAGE | BEAPATHY | BEDEP | BEVHALL |
| BEAHALL | BEDEL | BEDISIN | BEIRRIT | BEAGIT |
| BEPERCH | BEOTHR | NACCBEHF | BEMODE | MOGAIT |
| MOFALLS | MOTREM | MOSLOW | NACCMOTF | MOMODE |
| COURSE | FRSTCHG | MMSEORDA | MMSEORLO | NACCMMSE |
| LOGIMEM | MEMUNITS | MEMTIME | DIGIF | DIGIFLEN |
| DIGIB | DIGIBLEN | ANIMALS | VEG | TRAILA |
| TRAILB | WAIS | BOSTON | COGSTAT | NORMCOG |
| DEMENTED | NACCPPA | NACCPPME | NACCBVFT | NACCLBDS |
| NACCTMCI | NACCMCIL | NACCMCIA | NACCMCIE | NACCMCIV |
| IMPNOMCI | NACCALZD | NACCALZP | PROBAD | PROBADIF |
| POSSAD | POSSADIF | NACCLBDE | NACCLBDP | PARK |
| PSP | PSPIF | CORT | CORTIF | FTD |
| FTDIF | PPAPH | PPAPHIF | VASC | VASCIF |
| VASCPS | VASCPSIF | STROKE | STROKIF | DOWNS |
| DOWNSIF | HUNT | HUNTIF | PRION | PRIONIF |
| BRNINJ | BRNINJIF | HYCEPH | HYCEPHIF | NEOP |
| NEOPIF | DEP | DEPIF | OTHPSY | OTHPSYIF |
| ALCDEM | ALCDEMIF | DYSILL | DYSILLIF | MEDS |
| MEDSIF | DEMUN | DEMUNIF | NACCETPR | NACCADMU |
| NACCFTDM | NACCAGE | NACCAAAS | NACCAANX | NACCAC |
| NACCACEI | NACCADEP | NACCADMD | NACCAHTN | NACCAMD |
| NACCANGI | NACCAPSY | NACCBETA | NACCCCBS | NACCDBMD |
| NACCDIUR | NACCEMD | NACCEPMD | NACCHTNC | NACCLIPL |
| NACCNSD | NACCPDMD | NACCVASD | NACCBMI | NACCUDSD |
| NACCFTD | NACCLBDM | NACCAPOE | NACCNE4S | NRDAYS |
| HEARFUNC | VISFUNC | |  |  |

**S.5 A screenshot of our Shiny App**

Chart

Description automatically generated with low confidence

**Figure S1 Screenshot of the model Shiny App**