


RESEARCH ARTICLE

Developing and validating a prediction tool for cerebral amyloid angiopathy neuropathological severity

Chenyin Chu^{1,2} | Yihan Wang^{1,2} | Liwei Ma^{1,2} | Chloe A. Mutimer³ | Guangyan Ji¹ | Huiyu Shi¹ | Nawaf Yassi^{3,4} | Colin L. Masters¹ | Benjamin Goudey^{2,5} | Liang Jin^{1,2} | Yijun Pan^{1,2} ¹The Florey Institute of Neuroscience and Mental Health, Parkville, Victoria, Australia²Florey Department of Neuroscience and Mental Health, The University of Melbourne, Parkville, Victoria, Australia³Department of Medicine and Neurology, Melbourne Brain Centre at The Royal Melbourne Hospital, The University of Melbourne, Parkville, Victoria, Australia⁴Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia⁵The ARC Training Centre in Cognitive Computing for Medical Technologies, The University of Melbourne, Calton, Victoria, Australia

Correspondence

Yijun Pan and Liang Jin, The University of Melbourne, 30 Royal Parade, Parkville, Victoria 3052, Australia.

Email: yijun.pan@unimelb.edu.au; liang.jin@unimelb.edu.au

Liang Jin and Yijun Pan contributed as co-senior authors.

Funding information

National Health and Medical Research Council; Alzheimer's Association, Grant/Award Numbers: GNT2007912, GNT2022203

Abstract

INTRODUCTION: Cerebral amyloid angiopathy (CAA) is a cerebrovascular condition, the severity of which can only be determined *post mortem*. Here, we developed machine learning models, the Florey CAA Score (FCAAS), to predict CAA severity (none/mild/moderate/severe).**METHODS:** Building on an auto-score-ordinal algorithm, the FCAAS models were developed and validated using data collected by three cohort studies of aging and dementia. The developed FCAAS models were digitized as a web-based tool. A pilot trial was conducted using this web-based tool.**RESULTS:** The FCAAS-4 achieved a mean area under the receiver operating characteristic curve (AUC-ROC) of 0.74 (95% confidence interval: 0.71–0.77) and a Harrell generalized c-index of 0.72 (0.70–0.75). Pilot trial results obtained a mean AUC-ROC of 0.82 (0.71–0.85) and Harrell generalized c-index 0.79 (0.73–0.82).**DISCUSSION:** The FCAAS models demonstrate a promising performance in predicting CAA severity. This framework holds the potential for predicting development of amyloid-related imaging abnormalities (ARIAs), given the CAA–ARIAs link.

KEYWORDS

amyloid-related imaging abnormalities, AutoScore ordinal algorithm, cerebral amyloid angiopathy, machine learning, risk prediction

Highlights

- The severity of cerebral amyloid angiopathy (CAA) can only be determined *post mortem*.
- A web tool, the Florey CAA Score (FCAAS), was developed to predict CAA severity.
- The FCAAS holds the potential to be used for CAA risk stratification in clinics.
- CAA is linked to increased risk of amyloid-related imaging abnormalities (ARIAs).
- The framework used by FCAAS can possibly be adapted to predict ARIAs risk.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Alzheimer's & Dementia* published by Wiley Periodicals LLC on behalf of Alzheimer's Association.

1 | BACKGROUND

Cerebral amyloid angiopathy (CAA) is a cerebrovascular condition characterized by the accumulation of amyloidogenic proteins in the cerebral blood vessel walls, leading to a weakened vasculature with increased risk of intracerebral hemorrhages.¹ Recently, CAA has been brought to the attention of more clinicians and dementia researchers due to its link to the immunotherapy-related adverse effect in Alzheimer's disease (AD). The adverse effect, amyloid-related imaging abnormalities (ARIA), has been observed in the clinical trials of monoclonal antibody drugs (mAbs) targeting amyloid-beta (A β) in AD patients.^{2,3}

The diagnosis of CAA is not straightforward. As per the Boston 2.0 Criteria,⁴ probable or possible CAA can be diagnosed clinically in patients aged ≥ 50 years who have characteristic hemorrhagic findings and/or white matter features on brain magnetic resonance imaging (MRI) scan in the absence of an alternative cause.⁵ However, a definitive diagnosis of CAA is only achievable through *post mortem* examination of the brain,⁶ which is also currently the only approach to determine the severity of CAA (none, mild, moderate, severe).⁷ Tools that can predict the severity of CAA would be clinically important. These tools have a wider use, with the potential to inform the use of antithrombotics in people with CAA; for example, it can prompt a risk-benefit discussion with patients predicted to develop severe CAA. In addition, given the link between CAA and ARIAs,^{8,9} these tools can possibly be adapted to facilitate ARIAs risk stratification, which can also inform ARIAs MRI monitoring protocols.

Here, we developed and validated four machine learning models, the Florey CAA Score (FCAAS). These models were built on an AutoScore ordinal algorithm and leveraged on data collected by three cohort studies of aging and dementia. All models were designed for classifying CAA severity, each addressing different ranges and extents of severity to provide flexibility in clinical applications. Importantly, the FCAAS has achieved promising performance in a simulation pilot trial.

2 | METHODS

2.1 | Data sources and ethics

In this study, we use data from the Religious Orders Study and the Memory and Aging Project (ROSMAP)¹⁰ for model development and validation, and data from the Latino CORE Study (LATC)¹¹ and the Minority Aging Research Study (MARS)¹² were used for simulation pilot trials. These studies use similar data collection protocols and inclusion/exclusion criteria, except that their targeted ethnicity populations are different. Data can be requested from <https://www.radc.rush.edu/>. Each study was approved by the institutional review board of Rush University Medical Center. All participants provided informed consent, signed an Anatomic Gift Act for brain donation, and agreed to a repository agreement, allowing their data and biospecimens to be shared. Data collection for these longitudinal studies has been described in previous studies.^{10–12} The present study performed

RESEARCH IN CONTEXT

1. **Systematic review:** The authors conducted a literature search using PubMed to explore diagnostic methods for cerebral amyloid angiopathy (CAA). The findings revealed that probable CAA is typically diagnosed via magnetic resonance imaging, while definitive diagnosis requires *post mortem* neuropathology. Currently, no tool exists to predict CAA neuropathological severity.
2. **Interpretation:** We developed and validated the Florey CAA Score (FCAAS) to predict CAA neuropathological severity. This model was digitized as a web-based tool, and its pilot trial demonstrated promising performance. In addition, our machine learning feature selection process identified key risk factors that may potentially contribute to more severe CAA.
3. **Future directions:** The FCAAS holds the potential to predict intracerebral hemorrhage risk, contingent upon the predicted CAA severity being evaluated against clinical outcomes in future studies. Given that more severe CAA is associated with a higher risk of amyloid-related imaging abnormalities (ARIA), the framework used by FCAAS could also be explored for predicting ARIAs risk.

secondary data analysis, which is exempted from ethics approval requirements.

2.2 | Participants and features

The study initially recruited 2118 participants from the ROSMAP, LATC, and MARS studies who have brain *post mortem* data available. For pre-processing, variables with $> 10\%$ missing data were removed from further analysis. The variables included in the study are listed in Table S1 in supporting information, and fall under nine categories, including clinical diagnosis, cognitive function, demographics, depression, disability, genetics, lifestyle, medical conditions, and neuropathology. For neuropathology data, we retained only amyloid density and tangle density, as they can be assessed using positron emission tomography imaging in living patients. This is an important consideration in preparing the FCAAS for practical clinical application. In addition, we removed participants who have missing data for any of the variables included. This process resulted in a final selection of 1727 ROSMAP and 60 LATC/MARS participants in the current study. The prediction outcome is the CAA severity (none, mild, moderate, severe).

2.3 | FCAAS model construction

The severity of CAA pathology follows a natural order (none, mild, moderate, severe), and therefore is an ordinal outcome. The development of the FCAAS model is based on an algorithm and R package

called AutoScore ordinal,¹³ an automated framework designed to facilitate the development and validation of risk prediction models for ordinal outcomes by systematically identifying potential features from high-dimensional data. An architecture framework has been prepared to illustrate the model construction process (Figure S1 in supporting information).

The AutoScore ordinal algorithm comprises six modules: (1) Feature ranking: all variables are ranked based on their importance using a random forest algorithm for ordinal classification. (2) Variable transformation: all continuous variables (age, alcohol consumption, amyloid density, blood pressure, body mass index, Mini-Mental State Examination (MMSE) score, social network size, tangle density, years of education) are converted to categorical variables to simplify interpretation and account for potential non-linear relationships between the features and the outcome. This approach is widely used for medical data to reduce the impact of outliers on model performance.¹⁴ (3) Score derivation: weights associated with variables are developed using the cumulative link model, a widely used regression model in studies of ordinal outcomes.¹⁵ Scores for each variable are generated through rescaling and normalization of the weights. (4) Feature number determination: this step involves a parsimony analysis with a 10-fold cross-validation to balance model performance and complexity, ultimately determining the final number of features for the FCAAS. (5) Fine-tuning: the cut-off values for continuous variables generated by the algorithm may not be clinically meaningful, and therefore they are fine-tuned with cut-offs adjusted based on standard clinical norms (e.g., clinical guidelines). (6) Model evaluation: the developed model is evaluated on the unseen dataset. Details of the AutoScore ordinal framework and data handling are available in Methods S1 in supporting information. In this study, the ROSMAP dataset was randomly divided into a non-overlapping training set (70%) and a test set (30%). The training set was used for training, development, and fine-tuning of the FCAAS (module 1–5), and the test set served as unseen data to evaluate the performance (module 6).

We developed the FCAAS models by considering (1) four categories (none, mild, moderate, severe; referred to as FCAAS-4), (2) three categories (none, mild, moderate/severe; referred to as FCAAS-3), and (3) two categories (none, mild/moderate/severe, referred to as FCAAS-2a; none/mild, moderate/severe, referred to as FCAAS-2b). These models were designed to meet potential needs of clinical application; for example, FCAAS-4 can predict disease severity, FCAAS-3 can differentiate mild and moderate/severe CAA from people without CAA pathology, while FCAAS-2a can predict if an individual would develop CAA or not. FCAAS-2b was designed to identify people with moderate/severe CAA, as they may have a much higher risk of ARIAs if they were provided the anti-A β mAbs. Developing multiple models will enable future studies to identify which risk stratification approach is most beneficial.

2.4 | FCAAS model evaluation

As illustrated in Figure S1, model evaluation was conducted on the test set using a mean area under the receiver operating characteris-

tic curve (AUC-ROC) and a Harrell generalized c-index as metrics for evaluation. For classification with N categories, it decomposes into $N-1$ binary classifications, and the AUC-ROC across these binary classifications is referred to as the mean AUC-ROC.¹⁶ The Harrell generalized c-index measures the proportion of concordant pairs (i.e., when predictions and observed outcomes generate the same ranking for a pair of observations, including tied ranks) among all possible pairs of observations. This is commonly used to evaluate classification models.^{17,18} For both the mean AUC-ROC and the Harrell generalized c-index, a value of 0.5 indicates random performance, while a value of 1 indicates perfect prediction performance. Evaluation outcomes were reported with a bias-corrected 95% bootstrap confidence interval (CI). Sensitivity and specificity were calculated for FCAAS-2a&2b, given they are for a binary classification.

The performance of the FCAAS models was compared to the following baseline models: a proportional odds model with least absolute shrinkage and selection operator feature selection (POM-1), a proportional odds model with stepwise feature selection (POM-2), and a proportional odds model using the features selected in modules 1 and 4 of FCAAS (POM-3). Comparing POM-1, POM-2, and POM-3 to FCAAS-3 and FCAAS-4 models highlights the advantages of the AutoScore ordinal algorithm over commonly used ordinal outcome classification models. For binary classification, logistic regression using all features (logistic-all) and logistic regression with random forest feature selection (logistic-selected) were also used as baseline models. Both models are frequently used for binary classification and were therefore selected for comparison to FCAAS-2a&2b.

2.5 | Web-based tool development

The FCAAS models developed in this study have been digitized as a web-based tool through a co-design process with health consumers and clinicians.¹⁹ This is to ensure the web-based tool is more likely to be accepted by the end users after its implementation. The co-design was held virtually and physically via the Victorian Co-design Research Hub. This tool prompts clinicians to input data into several specifically chosen features. Once the required information is entered, the tool calculates the FCAAS at the web backend and the predicted CAA pathology will be made available to the clinicians. In this paper, we only showcase the FCAAS-4 to avoid repetition.

2.6 | Simulation pilot trial of FCAAS-4 for LATC and MARS

To test the FCAAS web-based tool, a simulation pilot trial with 60 participants drawn from the LATC and MARS studies was conducted. These data were not seen during the FCAAS model construction and validation. The FCAAS results were computed using the AutoScore ordinal algorithm, handled by a researcher (G.J.) who was blind to the participants' clinical diagnoses. The predicted results from the pilot

trial were then compared to the clinical diagnoses to assess mean AUC-ROC and Harrell generalized c-index.

2.7 | Software and packages

All data preprocessing and analyses were performed using Python version 3.9 and RStudio version 12.0+369. The developed model relies on the AutoScore ordinal package in the R 3.5.3 programming environment (R Foundation).²⁰ The AutoScore ordinal package enables the convenient creation of point-based clinical scoring models to predict outcomes, minimizing manual intervention for data processing, parameter tuning, and model optimization.

3 | RESULTS

3.1 | Participant characteristics

Using an AutoScore ordinal algorithm, the FCAAS model was developed based on 1727 participants in the ROSMAP study, and a pilot trial on 60 LATC and MARS participants. For ROSMAP participants, the mean age at baseline was 80.10 years (standard deviation [SD] = 7.00). The average MMSE score at baseline was 27.60 (SD = 2.83). Approximately 68.9% of the participants are female, with a mean education level of 16.29 years (SD = 3.58). At death, 21%, 41%, 23%, and 15% participants were diagnosed neuropathologically with none, mild, moderate, or severe CAA, respectively. For LATC and MARS, the mean age at baseline was 75.93 years (SD = 6.65). The average MMSE score at baseline was 27.13 (SD = 3.31). Approximately 66.7% of the participants are female, with a mean education level of 14.48 years (SD = 3.13). At death, 43%, 18%, 25%, and 14% participants were diagnosed neuropathologically with none, mild, moderate, or severe CAA, respectively. The detailed participant characteristics are summarized in Table 1 and Table S2 in supporting information.

3.2 | Outcomes of FCAAS-4 model construction and evaluation

We first performed a random forest feature selection to identify the most to least important features for ordinal classification. The feature importance ranking (Figure 1) highlights that amyloid density and tangle density are the two most important features related to CAA. This is consistent with the observation that higher A β levels are a crucial prerequisite for vascular amyloidosis,²¹ and that CAA is significantly associated with the formation of neurofibrillary tangles.²² The third most important feature is age, which is consistent with the clinical observation that age is a risk factor for CAA.⁷ Body mass index (BMI) also emerges as a key feature, although this conflicts with a previous finding showing no significant difference in BMI between people with and without CAA.²³ In addition, variables such as blood pressure, social network size (the sum of all individuals with whom an individual has contact at least once every 2 weeks), education, apolipoprotein E

(APOE) genotype, and MMSE score are highlighted as important features, which are generally supported by clinical and epidemiological evidence.^{24–26}

To balance the complexity and performance of the FCAAS-4 model, a parsimony analysis was conducted to determine the optimal number of features. As shown in the parsimony plot (Figure 2), we observed a sharp increase in the mean AUC-ROC when the FCAAS model incorporated one to five features (AUC-ROC 0.45–0.68). The mean AUC-ROC gradually increased when more features were added, reaching its peak at 10 features (mean AUC-ROC 0.68–0.75). Additional features did not improve the mean AUC-ROC further. Therefore, 10 features were used by the FCAAS-4 model, balancing model complexity and performance.

After feature selection, a cumulative log-link regression model was used to derive scores for each category/interval of the selected features (Table 2, left panel). The maximum FCAAS score is 100. The performance metrics on the validation set were 0.76 (95% CI, 0.75–0.78) for mean AUC-ROC and 0.73 (0.72–0.75) for the Harrell generalized c-index. We thereafter manually fine-tuned the intervals for clinical relevance. The refined intervals were [0, 0.72), [0.72, 0.93), [0.93, 1.28), and ≥ 1.28 for amyloid density;²⁷ [0, 0.79), [0.79, 1.11), [1.11, 1.91), and ≥ 1.91 for tangle density;²⁷ [0, 18.5), [18.5, 25), [25, 30), and ≥ 30 for BMI.²⁸ For diastolic blood pressure, the intervals are < 60, [60, 80), [80, 90), and ≥ 90 , while for systolic blood pressure, the intervals are < 120, [120, 150), and ≥ 150 .²⁹ For MMSE, the intervals are [0, 25), [25, 29).^{28–30} The updated scores for the selected features, after fine-tuning, are presented in Table 2, right panel.

The performance metrics of the fine-tuned FCAAS-4 model on the validation set were 0.75 (95% CI, 0.73–0.77) for mean AUC-ROC, and 0.73 (0.71–0.74) for the Harrell generalized c-index, which were comparable to the performance of the FCAAS-4 model without fine-tuning. Despite this, incorporating fine-tuning is necessary, as it integrates clinical information into the score derivation process, allowing for a more meaningful interpretation compared to the quantile-based intervals determined by the Auto-Score algorithm.

After fine-tuning, the FCAAS-4 was evaluated on the test set, achieving a mean AUC-ROC of 0.73 (95% CI, 0.72–0.74) and a Harrell generalized c-index of 0.71 (0.70–0.72). The score was mapped to the predicted CAA severity, as per Table 3, which was generated by the AutoScore algorithm. The score of none, mild, moderate, and severe CAA fall in the range of [0, 20), [20, 50), [50, 55), and [55, 100], respectively. We also plotted the risk for non-overlapping intervals to illustrate the probability of different CAA severity in Figure 3, which demonstrates that as the score increases, the probability of more severe CAA increases. Overall, these results indicate that the FCAAS-4 model can distinguish people with CAA pathology from those without and differentiate CAA severity.

3.3 | Outcomes of FCAAS-3 and FCAAS-2a&2b model construction and evaluation

Additional models were developed to classify CAA across different ranges or extents of severity, potentially addressing varying clinical

TABLE 1 Participant characteristics.

Features	ROSMAP	LATC and MARS
Cognitive status	CU: 1160; MCI: 451; AD: 116	CU: 43; MCI: 12; AD: 5
Self-reported memory complaints	Yes: 1113; No: 614	Yes: 22; No: 38
Mini-Mental State Examination (MMSE)	27.60 (2.83)	27.13 (3.31)
Age (years)	80.10 (7.00)	75.93 (6.65)
Education (years)	16.29 (3.58)	14.48 (3.13)
Sex	Female: 1190, Male: 537	Female: 40, Male: 20
Modified epidemiologic studies depression scale	1.14 (1.56)	0.98 (1.14)
Instrumental activities	1.13 (1.58)	0.62 (0.92)
Basic activities	0.17 (0.61)	0.10 (0.44)
Mobility disability	0.80 (0.97)	0.72 (0.96)
APOE genotype	$\epsilon 3/\epsilon 3$:1061; $\epsilon 4/\epsilon 3$:382; $\epsilon 3/\epsilon 2$:211; $\epsilon 4/\epsilon 4$:30; $\epsilon 4/\epsilon 2$:33; $\epsilon 2/\epsilon 2$:10	$\epsilon 3/\epsilon 3$:27; $\epsilon 4/\epsilon 3$:15; $\epsilon 3/\epsilon 2$:10; $\epsilon 4/\epsilon 4$:2; $\epsilon 4/\epsilon 2$:6; $\epsilon 2/\epsilon 2$:0
Body mass index	26.77 (4.89)	30.17 (7.68)
Daily alcohol intake (number of drink/day)	0.45 (0.98)	0.40 (0.99)
Smoking habits	Yes:1209; No:518	Yes:38; No:22
Social network size ^a	7.45 (8.52)	6.90 (9.46)
Diastolic blood pressure (mmHg)	72.62 (10.75)	76.79 (12.46)
Systolic blood pressure (mmHg)	135.71 (17.81)	135.53 (20.32)
Hypertension	Yes: 814; No: 913	Yes: 45; No: 15
Self-reported medical history ^b	1.37 (1.08)	1.75 (1.02)
Diabetes	Yes: 211; No: 1516	Yes: 17; No: 43
Heart conditions ^c	Yes: 214; No: 1513	Yes: 8; No: 52
Cancer	Yes: 569; No: 1158	Yes: 17; No: 43
Thyroid conditions	Yes 314; No: 1413	Yes 9; No: 51
Claudication	Yes: 111; No: 1616	Yes 6; No: 54
Vascular disease burden ^d	0.27 (0.53)	0.32 (0.62)
Vascular disease risk factor ^e	0.89 (0.78)	1.67 (0.90)
Amyloid density	1.05 (0.74)	0.78 (0.71)
Tangle density	1.35 (0.98)	1.19 (1.14)
CAA pathology ^f	0: 367; 1: 708; 2: 398; 3: 254	0: 26; 1: 11; 2: 15; 3: 8
Follow up years	8.95 (5.81)	8.72 (4.62)

Note: All data were collected at baseline except neuropathology data (i.e., amyloid density, tangle density, CAA pathology). Data were presented as mean (SD). Follow-up year was the time between baseline and participant deceased.

Abbreviations: AD, Alzheimer's disease; APOE, apolipoprotein E; CAA, cerebral amyloid angiopathy; CU, cognitively unimpaired; LATC, Latino CORE Study; MARS, Minority Aging Research Study; MCI, mild cognitive impairment; ROSMAP, Religious Orders Study and the Memory and Aging Project; SD, standard deviation.

^aNumber of community members, relatives, and friends seen at least once a month.

^bMeasure of seven medical conditions, including hypertension, diabetes, heart conditions, cancer, thyroid conditions, head injury, and stroke.

^cHeart attack, coronary thrombosis, coronary occlusion, myocardial infarction.

^dClaudication, stroke, heart conditions.

^eHypertension, diabetes, smoking history.

^f0 = none, 1 = mild, 2 = moderate, 3 = severe; four brain regions (midfrontal, midtemporal, parietal, and calcarine cortices) were assessed for CAA pathology in the meningeal and parenchymal vessels and the severity is determined by averaging all brain regions, semi-quantitatively assessed by neuropathologist.

needs. The development of FCAAS-3 and FCAAS-2a&2b followed the same methodology as FCAAS-4. In Module 1, random forest feature selection was used to rank features by their importance for FCAAS-3 and FCAAS-2a&2b, as shown in Figure S2 in supporting information. The variable rankings for these models were similar to the FCAAS-4,

with minor alterations in the importance values. The parsimony analysis for FCAAS-3 and FCAAS-2a&2b in Modules 2–4, along with the corresponding parsimony plots, are presented in Figure S3 in supporting information. According to the parsimony plots, the optimal number of features for FCAAS-3, FCAAS-2a, and FCAAS-2b is ten,

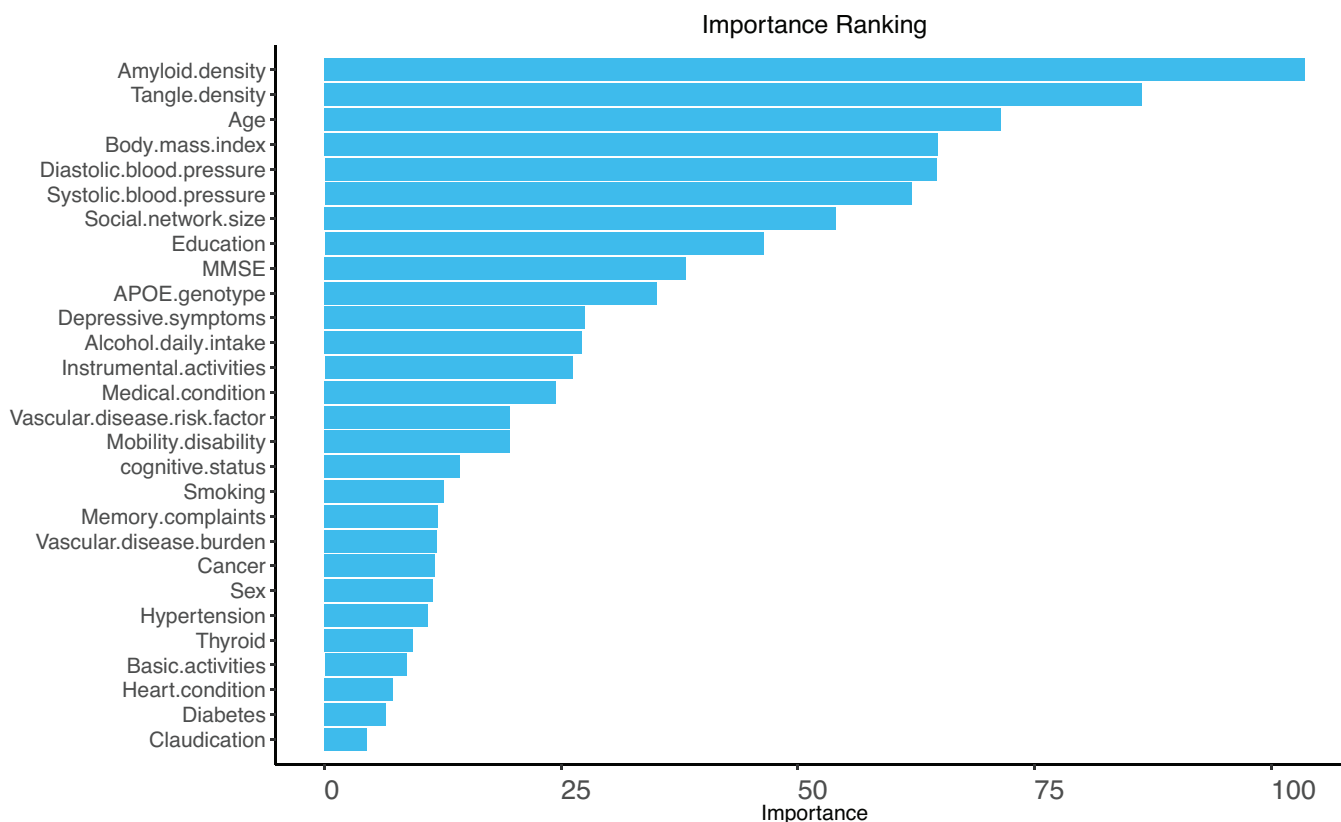


FIGURE 1 Feature ranking. The x axis displays the importance determined by the random forest feature selection, and the y axis lists the names of all the features included in the study. A wider bar indicates greater importance. APOE, apolipoprotein E; MMSE, Mini-Mental State Examination.

ten, and eight, respectively. The fine-tuned scores for FCAAS-3 and FCAAS-2a&2b are summarized in Table S3 in supporting information. The FCAAS-3 achieved a mean AUC-ROC of 0.76 (95% CI, 0.74–0.77) and a Harrell generalized c-index of 0.73 (0.72–0.75). The FCAAS-2a achieved an AUC-ROC of 0.78 (0.73–0.82), sensitivity of 0.71 (0.67–0.76), and specificity of 0.76 (0.69–0.84), with a threshold score of 41. The FCAAS-2b obtained an AUC-ROC of 0.76 (0.72–0.80), a sensitivity of 0.73 (0.68–0.79), and a specificity of 0.71 (0.66–0.75), with a threshold score of 35.

3.4 | Comparison between the baseline models and FCAAS

The comparisons among four FCAAS models and their baseline models have been summarized in Table 4. The performance of FCAAS-4 was better than POM-1 (mean AUC-ROC: 0.68 (95% CI, 0.66–0.69), Harrell generalized c-index: 0.60 (0.59–0.61)), POM-2 (mean AUC-ROC: 0.69 (0.67–0.70), Harrell generalized c-index: 0.64 (0.63–0.65)), and POM-3 (mean AUC-ROC: 0.69 (0.68–0.70), Harrell generalized c-index: 0.69 (0.67–0.70)). Similarly, FCAAS-3 demonstrated a superior performance compared to POM-1 (mean AUC-ROC: 0.72 (0.71–0.73), Harrell generalized c-index: 0.66 (0.65–0.67)), POM-2 (mean AUC-ROC: 0.70

(0.67–0.72), Harrell generalized c-index: 0.68 (0.67–0.69)), and POM-3 (mean AUC-ROC: 0.67 (0.65–0.69), Harrell generalized c-index: 0.69 (0.68–0.70)). The performance of FCAAS-2a was superior to the logistic-all model (AUC-ROC: 0.66 (0.61–0.71), sensitivity: 0.81 (0.74–0.84), specificity: 0.38 (0.32–0.46)), and the logistic-selected model (AUC-ROC: 0.70 (0.68–0.72), sensitivity: 0.83 (0.78–0.87), specificity: 0.38 (0.30–0.44)). The performance of FCAAS-2b was better than the logistic-all model (AUC-ROC: 0.65 (0.61–0.70), sensitivity: 0.75 (0.67–0.78), specificity: 0.39 (0.34–0.50)), and the logistic-selected model (AUC-ROC: 0.67 (0.63–0.72), sensitivity: 0.77 (0.71–0.81), specificity: 0.40 (0.31–0.49)).

3.5 | Simulation pilot trial of FCAAS-4

We have digitized the FCAAS-4 as a web-based tool, which automatically calculates the score once the required information is entered (a demo video is provided as [Supplemental video file](#) in supporting information). The calculated FCAAS scores for the trial participants are presented in Table S4 in supporting information. This trial achieved a mean AUC-ROC of 0.82 (0.71–0.85) and a Harrell generalized c-index of 0.79 (0.73–0.82). Overall, the pilot trial results support the potential clinical application of FCAAS in the CAA severity prediction.

Final Parsimony plot based on 10-fold Cross Validation

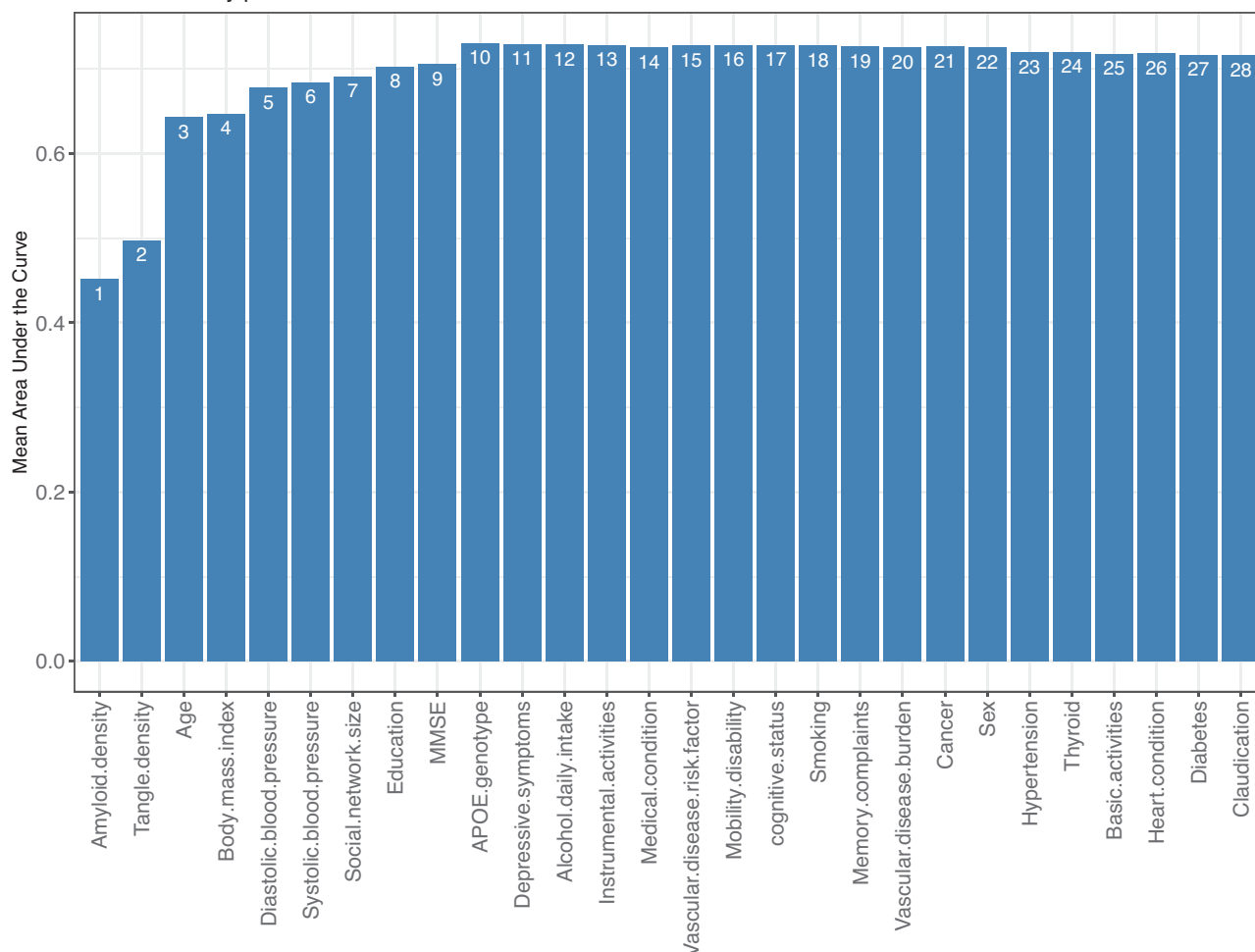


FIGURE 2 Parsimony plot for the FCAAS-4. The plot displays the average mean AUC-ROC values as an increasing number of features are used by the FCAAS-4 model. The number annotated on each bar represents the cumulative number of features used. Taller bars indicate better performance. APOE, apolipoprotein E; AUC-ROC, area under the receiver operating characteristic curve; FCAAS-4, Florey Cerebral Amyloid Angiopathy Score, 4-category model; MMSE, Mini-Mental State Examination.

4 | DISCUSSION

To our knowledge, this is the first report of machine learning models designed to predict CAA severity. The AutoScore ordinal algorithm ranked the importance of all predictive features and used a parsimony analysis to determine the optimal number of features required by the FCAAS, balancing model complexity and performance. In addition, a score table was generated, assigning different scores to each category of the selected features. From the score tables (Table 2, Table S3), we observed that APOE variants ($\epsilon 4/\epsilon 2$, $\epsilon 4/\epsilon 3$, $\epsilon 4/\epsilon 4$, and $\epsilon 2/\epsilon 2$), amyloid density (> 0.72), and tangle density (> 1.91) are major contributors to the FCAAS score, with the exception that APOE was not selected by the FCAAS-2a model. This observation is well aligned with clinical/epidemiological evidence.^{21,22} Of note, the APOE $\epsilon 4/\epsilon 4$ variants received the highest score among all the variants, which is likely due to the higher A β load associated with APOE $\epsilon 4$.³¹ We noted that APOE $\epsilon 2$ may also contributed to a higher score, which agrees with a previous study reporting both APOE $\epsilon 2$ and APOE $\epsilon 4$

being correlated with an increased amyloid accumulation in cerebral vasculature.³² Interestingly, as demonstrated in the score tables, hypertension and hypotension can both contribute to a higher score for FCAAS, which should be investigated in future studies. The FCAAS models achieved promising performances; the observation that all FCAAS models perform better than the baseline models (Table 4) highlights the advantages of using the Auto-Score ordinal algorithm for ordinal classification compared to the common algorithms used by the baseline models. Overall, these results highlight that our developed models are aligned with evidence-based medicine, and our selected algorithm is appropriate for the prediction of CAA severity.

To evaluate the performance of FCAAS on external dataset, we performed a simulation pilot trial on the LATC and MARS participants, and a mean AUC-ROC of 0.82 (0.71–0.85) and a Harrell generalized c-index of 0.79 (0.73–0.82) were achieved. This achievement is promising, especially considering the distinct ethnic backgrounds among the LATC, MARS, and ROSMAP participants. The FCAAS model demonstrates potential for generalizability to a diverse population. However,

TABLE 2 Score table for FCAAS-4 model.

Features	Before fine-tuning		After fine-tuning	
	Interval	Score	Interval	Score
Amyloid density*	[0, 0.151)	0	[0, 0.72)	0
	[0.151, 1.71)	16	[0.72, 0.93)	13
	[1.71, 2.18)	17	≥ 0.93	15
	≥ 2.18	18		
Tangle density*	[0, 0.253)	0	[0, 0.79)	0
	[0.253, 0.576)	2	[0.79, 1.91)	4
	[0.576, 1.99)	6	≥ 1.91	10
	[1.99, 3.31)	11		
	≥ 3.31	15		
Age	[0, 67.8)	1	[0, 67.8)	4
	[67.8, 73.8)	0	[67.8, 73.8)	0
	[73.8, 86.1)	2	[73.8, 86.1)	2
	[86.1, 90.6)	5	[86.1, 90.6)	4
	≥ 90.6	8	≥ 90.6	9
BMI*	[0, 19.8)	2	[0, 18.5)	0
	[19.8, 22.8)	5	[18.5, 25)	4
	[22.8, 30.5)	6	[25, 30)	5
	[30.5, 36.5)	5	≥ 30	3
	≥ 36.5	2		
Diastolic blood pressure*	[0, 55.2)	6	[0, 60)	4
	[55.2, 63)	0	≥ 60	0
	[63, 81)	1		
	[81, 90)	0		
	≥ 90	2		
Systolic blood pressure*	[0, 110)	5	<120	4
	[110, 121)	3	[120, 150)	0
	[121, 150)	0	≥ 150	4
	[150, 166)	3		
	≥ 166	2		
Social network size	[0, 1)	7	[0, 1)	7
	[1, 3)	0	[1, 3)	0
	[3, 20)	3	[3, 11)	3
	≥ 20	7	[11, 20)	4
			≥ 20	8
Education	[0, 19)	0	[0, 12)	8
	[19, 22)	2	[12, 19)	0
	≥ 22	0	[19, 22)	2
			≥ 22	0
APOE genotype	ε2/ε2	4	ε2/ε2	1
	ε2/ε3	4	ε2/ε3	5
	ε2/ε4	14	ε2/ε4	17
	ε3/ε3	0	ε3/ε3	0
	ε3/ε4	9	ε3/ε4	10
	ε4/ε4	28	ε4/ε4	33

(Continues)

TABLE 2 (Continued)

Features	Before fine-tuning		After fine-tuning	
	Interval	Score	Interval	Score
MMSE*	[0, 22)	1	[0, 25)	0
	[22, 30)	0	[25, 29)	2
			[29, 30)	1

Note: *fine-tuning based on clinical norms.

Abbreviations: APOE, apolipoprotein E; BMI, body mass index; FCAAS-4, Florey Cerebral Amyloid Angiopathy Score, 4-category model; MMSE, Mini-Mental State Examination.

TABLE 3 FCAAS-4 scoring with probability.

Score	None	Mild	Moderate	Severe
[0,5]	77.66%	19.22%	2.36%	0.75%
(5,10]	69.21%	26.05%	3.58%	1.16%
(10,15]	60.15%	32.94%	5.18%	1.72%
(15,20]	50.35%	39.7%	7.41%	2.54%
(20,25]	40.52%	45.35%	10.39%	3.74%
(25,30]	31.4%	48.92%	14.21%	5.47%
(30,35]	23.51%	49.75%	18.79%	7.94%
(35,40]	17.11%	47.69%	23.81%	11.39%
(40,45]	12.17%	43.12%	28.64%	16.07%
(45,50]	8.51%	36.87%	32.43%	22.19%
(50,55]	5.88%	29.94%	34.38%	29.8%
(55,60]	4.02%	23.24%	34.01%	38.72%
(60,65]	2.74%	17.38%	31.41%	48.47%
(65,70]	1.85%	12.61%	27.2%	58.34%
(70,100]	0.58%	4.29%	12.18%	82.95%

Abbreviation: FCAAS-4, Florey Cerebral Amyloid Angiopathy Score, 4-category model.

we must acknowledge that all datasets used in this study were collected by the Rush University Medical Center, and therefore they are strictly not completely independent, as the LATC/MARS cohorts may share data collection protocols with the ROSMAP cohort. Regardless, we do not expect the performance of FCAAS would be significantly affected by dataset bias, given the use of Auto-Score algorithm using a scoring system. We have previously constructed and validated a different model, the Florey Dementia Risk Score (FDRS), using the Auto-Score algorithm for binary classification of AD dementia (healthy vs. diseased).³³ The Auto-Score ordinal algorithm used in the current study is an advanced version extending the binary classification to the ordinal outcome prediction. For the FDRS, an AUC of 0.82 (95% CI, 0.75–0.88) was achieved for internal validation with the Australian Imaging, Biomarker and Lifestyle (AIBL) dataset, and an AUC of 0.82 (95% CI, 0.77–0.86) was achieved when the FDRS was externally evaluated using the ROSMAP data.

The FCAAS tools may have several potential clinical applications. This prediction tool could be used to inform antithrombotic use in people with CAA as mentioned earlier. With further validation, this tool may offer the possibility to predict intracerebral hemorrhage risk,

if the predicted CAA severity can be evaluated against clinical outcomes in future studies. Although there is a possible link between CAA and ARIAs risk from receiving mAbs,^{9,34,35} whether the FCAAS can be used for ARIAs risk stratification is yet to be explored. It must be acknowledged that translating *post mortem* measurements into in vivo assessments can be challenging, and therefore the FCAAS models will require further optimization. Regardless, the AutoScore ordinal algorithm, used by the FCAAS, may have the potential for developing new models to predict ARIAs. Future studies exploring risk factors and biomarkers for ARIAs for the development of such models are needed.

ARIAs are currently detected through brain MRI scans. To monitor the occurrence of ARIAs, an MRI scan is recommended prior to the 5th, 7th, and 14th infusions of lecanemab. An additional MRI scan prior to the 26th infusion is also recommended, especially for those who are APOE ε4 carriers or had evidence of ARIAs (even asymptomatic) on earlier MRIs.³⁶ Our global health systems, however, are not fully equipped to provide anti-Aβ mAbs as early interventions for people living with AD. This is partially due to the costly monitoring for ARIAs, which presents challenges not only in funding the necessary scans but also in accessing scanners within a reasonable distance for patients, and the availability of specialized radiologists to read the scans. From the patient perspective, the need for repetitive imaging would be an accumulative burden on older adults. Our FCAAS if adapted for ARIAs risk stratification can be a potential solution to this emerging issue in dementia treatment by informing the ARIAs monitoring protocol.

The performance of the FCAAS models may be influenced by several potential sources of error and bias. For example, some features were self-reported by the study participants (e.g., social network size), which could introduce recall bias. In addition, the model was trained, validated, and evaluated using datasets collected from related cohort studies (ROS, MAP, LATC, MARS). It would be ideal to externally evaluate our models using a completely independent dataset to ensure model robustness and generalizability; however, CAA pathology is not often collected in dementia cohort studies. Moreover, participants in the ROS study were older religious clergy from across the United States, although the MAP study expanded the cohort to include non-clergy participants from the Chicago area. This could contribute to selection bias. Finally, our predictive outcome (CAA severity) was based on *post mortem* data, which are limited by the absence of a reliable, definitive diagnostic method in living patients. Moreover, the measurement of amyloid density and tau density is currently only included in research settings, which is a limitation for immediate trans-

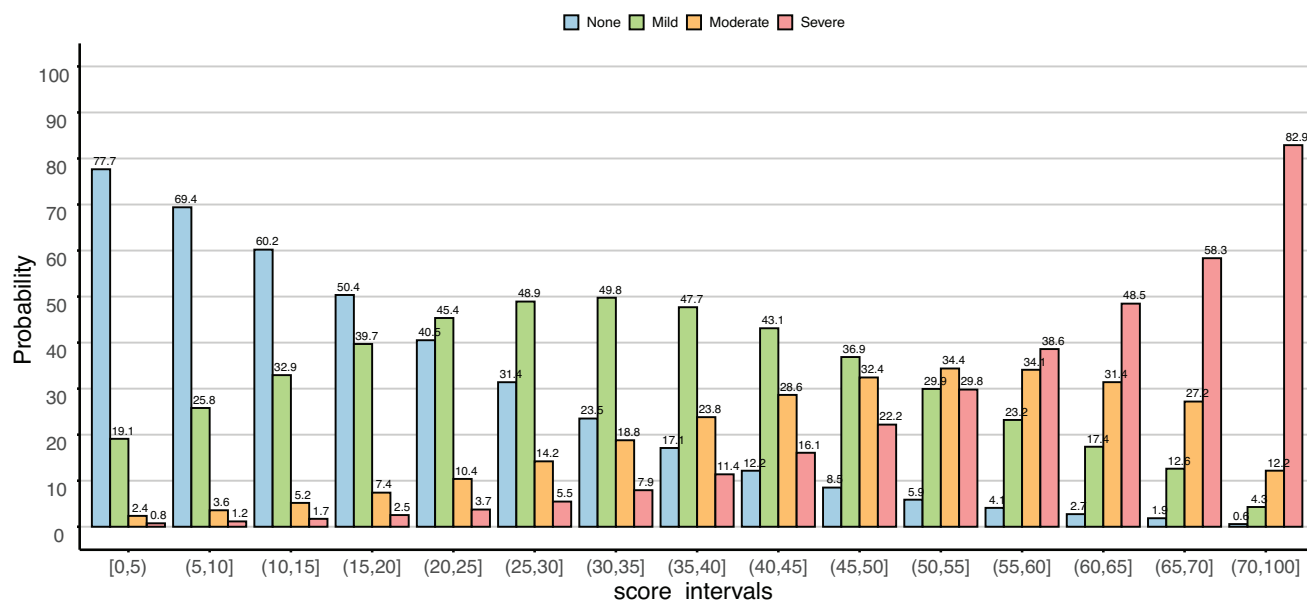


FIGURE 3 Score intervals and risk probabilities for the FCAAS-4 model. The x axis represents the FCAAS score intervals, and the y axis indicates the probability of CAA severity (color-coded). As the FCAAS score increases, the probability of more severe CAA also increases. CAA, cerebral amyloid angiopathy; FCAAS-4, Florey Cerebral Amyloid Angiopathy Score, 4-category model.

TABLE 4 Comparison between FCAAS models and baseline models.

Model	Mean AUC-ROC	Generalized c-index	AUC-ROC	Specificity	Sensitivity
FCAAS-4	0.73 (0.72–0.74)	0.71 (0.70–0.72)	NA	NA	NA
POM-1	0.68 (0.66–0.69), $p < 0.001$	0.60 (0.59–0.61), $p < 0.001$	NA	NA	NA
POM-2	0.69 (0.67–0.70), $p < 0.001$	0.64 (0.63–0.65), $p < 0.001$	NA	NA	NA
POM-3	0.69 (0.68–0.70), $p < 0.001$	0.69 (0.67–0.70), $p < 0.009$	NA	NA	NA
FCAAS-3	0.76 (0.74–0.77)	0.73 (0.72–0.75)	NA	NA	NA
POM-1	0.72 (0.71–0.73), $p = 0.002$	0.66 (0.65–0.67), $p = 0.001$	NA	NA	NA
POM-2	0.70 (0.67–0.72), $p = 0.001$	0.68 (0.67–0.69), $p = 0.001$	NA	NA	NA
POM-3	0.67 (0.65–0.69), $p = 0.001$	0.69 (0.68–0.70), $p = 0.002$	NA	NA	NA
FCAAS-2a	NA	NA	0.78 (0.73–0.82)	0.76 (0.69–0.84)	0.71 (0.67–0.76)
Logistic-all (a)	NA	NA	0.65 (0.61–0.70), $p < 0.001$	0.39 (0.34–0.50)	0.75 (0.67–0.78)
Logistic-selected (a)	NA	NA	0.67 (0.63–0.72), $p < 0.001$	0.40 (0.31–0.49)	0.77 (0.71–0.81)
FCAAS-2b	NA	NA	0.76 (0.72–0.80)	0.71 (0.66–0.75)	0.73 (0.68–0.79)
Logistic-all (b)	NA	NA	0.66 (0.61–0.71), $p < 0.001$	0.38 (0.32–0.46)	0.81 (0.74–0.84)
Logistic-selected (b)	NA	NA	0.70 (0.68–0.72), $p < 0.001$	0.38 (0.30–0.44)	0.83 (0.78–0.87)

Note: Purple = four classification (none, mild, moderate, severe); orange = three classification (none, mild, moderate/severe); blue = binary classification (none/mild, moderate/severe); green = binary classification (none/mild, moderate, severe). The AUC-ROC of each FCAAS model was compared to the baseline models using the Wilcoxon signed-rank test, and the P value for each comparison was annotated for the corresponding baseline.

Abbreviations: AUC-ROC, area under the receiver operating characteristic curve; FCAAS-4, Florey Cerebral Amyloid Angiopathy Score; POM, proportional odds model.

lation of FCAAS models into the clinics. However, the FCAAS models can be readily updated as new predictive outcomes from alternative diagnostic/quantification methods become available.

In conclusion, the FCAAS tools have been developed to predict the severity of CAA neuropathology. The score-based risk prediction system used by the FCAAS and its web-based platform can be used by clinicians without prior knowledge in machine learning. Several

potential clinical applications have been proposed; however, this would require careful clinical and implementation evaluation before this tool can be translated into clinical practice.

ACKNOWLEDGMENTS

This work would not have been possible without the contributions of the participants from ROSMAP, LATC, MARS studies, and their inves-

tigators and staff. These studies received support from the National Institute on Aging (grant numbers: P30AG10161, P30AG72975, R01AG15819, R01AG17917, U01AG46152, U01AG61356, P30AG072975 and R01AG22018). RADC resources can be requested at <https://www.radc.rush.edu> and www.synapse.org. The FCAAS model was digitized to a web-based tool by Shuowen Yu. The Victorian Co-design Research Hub was funded by the Telematics Trust. Dr. Yijun Pan is supported by the National Health and Medical Research Council, Australia (GNT2007912, GNT2022203) and Alzheimer's Association, USA (23AARF-1020292).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest. Author disclosures are available in the [supporting information](#).

CONSENT STATEMENT

Written informed consent was obtained from all participants by Rush University to allow secondary analysis of the collected data.

ORCID

Yijun Pan  <https://orcid.org/0000-0002-1442-3333>

REFERENCES

- Jäkel L, De Kort AM, Klijn CJ, Schreuder FH, Verbeek MM. Prevalence of cerebral amyloid angiopathy: a systematic review and meta-analysis. *Alzheimers Dement*. 2022;18:10-28.
- Sperling RA, Jack Jr CR, Black SE, et al. Amyloid-related imaging abnormalities in amyloid-modifying therapeutic trials: Recommendations from the Alzheimer's Association Research Roundtable Workgroup. *Alzheimers Dement*. 2011;7:367-385.
- Withington CG, Turner RS. Amyloid-related imaging abnormalities with anti-amyloid antibodies for the treatment of dementia due to Alzheimer's disease. *Front Neurol*. 2022;13:862369.
- Charidimou A, Boulouis G, Frosch MP, et al. The Boston criteria version 2.0 for cerebral amyloid angiopathy: a multicentre, retrospective, MRI-neuropathology diagnostic accuracy study. *Lancet Neurol*. 2022;21:714-725.
- Kozberg MG, Perosa V, Gurol ME, van Veluw SJ. A practical approach to the management of cerebral amyloid angiopathy. *Int J Stroke*. 2021;16:356-369.
- Kuhn J, Sharman T. Cerebral Amyloid Angiopathy. *StatPearls [Internet]*. StatPearls Publishing; 2024.
- Boyle PA, Yu L, Nag S, et al. Cerebral amyloid angiopathy and cognitive outcomes in community-based older persons. *Neurology*. 2015;85:1930-1936.
- Sperling R, Salloway S, Brooks DJ, et al. Amyloid-related imaging abnormalities in patients with Alzheimer's disease treated with bapineuzumab: a retrospective analysis. *Lancet*. 2012;11:241-249.
- Hampel H, Elhage A, Cho M, Apostolova LG, Nicoll JAR, Atri A. Amyloid-related imaging abnormalities (ARIA): radiological, biological and clinical characteristics. *Brain*. 2023;146:4414-4424.
- Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA. Religious Orders Study and Rush Memory and Aging Project. *J Alzheimers Dis*. 2018;64:S161-S189.
- Marquez DX, Glover CM, Lamar M, et al. Representation of older Latinxs in cohort studies at the Rush Alzheimer's Disease Center. *Neuroepidemiology*. 2020;54:404-418.
- Barnes LL, Shah RC, Aggarwal NT, Bennett DA, Schneider JA. The Minority Aging Research Study: Ongoing efforts to obtain brain donation in African Americans without dementia. *Curr Alzheimer Res*. 2012;9:734-745.
- Saffari SE, Ning Y, Xie F, et al. AutoScore-Ordinal: an interpretable machine learning framework for generating scoring models for ordinal outcomes. *BMC Med Res Methodol*. 2022;22:286.
- Bouwmeester W, Zuihthoff NP, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med*. 2012;9:e1001221.
- Christensen RHB. Cumulative link models for ordinal regression with the R package ordinal. *J Stat Softw*. 2018;35.
- Landgrebe TC, Duin RP. Approximating the multiclass ROC by pairwise analysis. *Pattern Recognit Lett*. 2007;28:1747-1758.
- Cheung LC. *Mixture models for left-and interval-censored data and concordance indices for composite survival outcomes*. The George Washington University; 2017.
- Brentnall AR, Cuzick J. Use of the concordance index for predictors of censored survival data. *Stat Methods Med Res*. 2018;27:2359-2373.
- Bird M, McGillion M, Chambers E, et al. A generative co-design framework for healthcare innovation: development and application of an end-user engagement framework. *Res Involv Engagem*. 2021;7:1-12.
- Xie F, Ning Y, Liu M, et al. A universal AutoScore framework to develop interpretable scoring systems for predicting common types of clinical outcomes. *STAR*. 2023;4:102302.
- Kumar-Singh S. Cerebral amyloid angiopathy: pathogenetic mechanisms and link to dense amyloid plaques. *Genes Brain Behav*. 2008;7:67-82.
- Attems J, Jellinger KA, Lintner F. Alzheimer's disease pathology influences severity and topographical distribution of cerebral amyloid angiopathy. *Acta Neuropathologica*. 2005;110:222-231.
- Zhang Q, Wang A, Meng X, et al. Vascular risk factors, imaging, and outcomes in transient ischemic attack/ischemic stroke patients with neuroimaging evidence of probable/possible cerebral amyloid angiopathy. *Oxid Med Cell Longev*. 2021;2021:995885.
- Ighodaro ET, Abner EL, Fardo DW, et al. Risk factors and global cognitive status related to brain arteriosclerosis in elderly individuals. *Blood Flow*. 2017;37:201-216.
- Yamada M, Naiki H. Cerebral amyloid angiopathy. *Prog Mol Biol Transl Sci*. 2012;107:41-78.
- Arvanitakis Z, Leurgans SE, Wang Z, Wilson RS, Bennett DA, Schneider JA. Cerebral amyloid angiopathy pathology and cognitive domains in older persons. *Ann Neurol*. 2011;69:320-327.
- Kapasi A, Poirier J, Hedayat A, et al. High-throughput digital quantification of Alzheimer disease pathology and associated infrastructure in large autopsy studies. *J Neuropathol Exp Neurol*. 2023;82:976-986.
- Weir CB, Jan A. BMI Classification Percentile And Cut Off Points. *StatPearls [Internet]*. StatPearls Publishing; 2023.
- Qaseem A, Wilt TJ, Rich R, et al. Pharmacologic treatment of hypertension in adults aged 60 years or older to higher versus lower blood pressure targets: A clinical practice guideline from the American College of Physicians and the American Academy of Family Physicians. *Ann Intern Med*. 2017;166:430-437.
- Shigemori K, Ohgi S, Okuyama E, Shimura T, Schneider E. The factorial structure of the mini mental state examination (MMSE) in Japanese dementia patients. *BMC*. 2010;10:1-7.
- Drzezga A, Grimmer T, Henriksen G, et al. Effect of APOE genotype on amyloid plaque load and gray matter volume in Alzheimer disease. *Neurology*. 2009;72:1487-1494.
- Nelson PT, Pious NM, Jicha GA, et al. APOE-ε2 and APOE-ε4 correlate with increased amyloid accumulation in cerebral vasculature. *J Neuropathol Exp Neurol*. 2013;72:708-715.
- Pan Y, Chu C, Wang Y, et al. Development and validation of the Florey Dementia Risk Score web-based tool to screen for Alzheimer's disease in primary care. *eClinicalMedicine*. 2024;76:102834.

34. Chantran Y, Capron J, Alamowitch S, Aucouturier P. Anti-A β antibodies and cerebral amyloid angiopathy complications. *Front Immunol*. 2019;10:1534.
35. Roytman M, Mashriqi F, Al-Tawil K, et al. Amyloid-related imaging abnormalities: an update. *Am J Roentgenol*. 2023;220:562-574.
36. Cummings J, Apostolova L, Rabinovici G, et al. Lecanemab: appropriate use recommendations. *J Prev Alzheimers Dis*. 2023; 10: 362-377.

How to cite this article: Chu C, Wang Y, Ma L, et al.

Developing and validating a prediction tool for cerebral amyloid angiopathy neuropathological severity. *Alzheimer's*

Dement. 2025;21:e14583. <https://doi.org/10.1002/alz.14583>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.