



OPEN

## Prediction of amyloid $\beta$ PET positivity using machine learning in patients with suspected cerebral amyloid angiopathy markers

Young Hee Jung<sup>1,2,3,8</sup>, Hyejoo Lee<sup>2,3,4,8</sup>, Hee Jin Kim<sup>2,3,4</sup>, Duk L. Na<sup>2,3,4,6,7</sup>, Hyun Jeong Han<sup>1</sup>, Hyemin Jang<sup>2,3,4</sup>✉ & Sang Won Seo<sup>2,3,4,5</sup>✉

Amyloid- $\beta$  (A $\beta$ ) PET positivity in patients with suspected cerebral amyloid angiopathy (CAA) MRI markers is predictive of a worse cognitive trajectory, and it provides insights into the underlying vascular pathology (CAA vs. hypertensive angiopathy) to facilitate prognostic prediction and appropriate treatment decisions. In this study, we applied two interpretable machine learning algorithms, gradient boosting machine (GBM) and random forest (RF), to predict A $\beta$  PET positivity in patients with CAA MRI markers. In the GBM algorithm, the number of lobar cerebral microbleeds (CMBs), deep CMBs, lacunes, CMBs in dentate nuclei, and age were ranked as the most influential to predict A $\beta$  positivity. In the RF algorithm, the absence of diabetes was additionally chosen. Cut-off values of the above variables predictive of A $\beta$  positivity were as follows: (1) the number of lobar CMBs > 16.4(GBM)/14.3(RF), (2) no deep CMBs(GBM/RF), (3) the number of lacunes > 7.4(GBM/RF), (4) age > 74.3(GBM)/64(RF), (5) no CMBs in dentate nucleus(GBM/RF). The classification performances based on the area under the receiver operating characteristic curve were 0.83 in GBM and 0.80 in RF. Our study demonstrates the utility of interpretable machine learning in the clinical setting by quantifying the relative importance and cutoff values of predictive variables for A $\beta$  positivity in patients with suspected CAA markers.

Cerebral amyloid angiopathy (CAA) is a cerebral small vessel disease (CSVD) characterized by amyloid  $\beta$  (A $\beta$ ) deposition in leptomeningeal and cortical vessels<sup>1,2</sup>. According to the modified Boston criteria, patients with multiple strictly lobar intracranial hemorrhage (ICH)/cerebral microbleeds (CMBs) or cortical superficial siderosis (cSS) on brain magnetic resonance imaging (MRI) are specific for CAA pathology, which leads to a clinicoradiological diagnosis of probable CAA<sup>3,4</sup>.

Recently, the clinical utility of A $\beta$  proton emission tomography (PET) in CAA patients has been widely investigated<sup>5-7</sup>. Based on previous evidence, A $\beta$  + PET scans in patients with CAA MRI markers may have clinical utility in two ways. First, A $\beta$  positivity in CAA patients enables clinicians to predict the prognosis of cognitive trajectories. Our previous study showed that A $\beta$  + patients with probable CAA had worse cognitive trajectories than their A $\beta$ - counterparts<sup>7</sup>. Several studies have emphasized the clinical significance of A $\beta$  + PET scans in assessing cognition in neurodegenerative diseases, including MCI<sup>8</sup>, AD, and vascular cognitive impairment<sup>9-11</sup>. Second, A $\beta$  PET positivity may provide insights into the underlying vascular pathology in patients with suspected CAA MRI markers; clinicians encounter patients with several lobar CMBs combined with a few deep CMBs who cannot be diagnosed as probable CAA based on criteria. However, these patients may have advanced CAA pathology, because CAA involvement propagates to deep areas in the later stage according to a pathologic study<sup>12</sup>. In this population, A $\beta$  positivity may suggest advanced CAA pathology rather than hypertensive angiopathy. This

<sup>1</sup>Department of Neurology, College of Medicine, Myoungji Hospital, Hanyang University, Goyang, Republic of Korea. <sup>2</sup>Department of Neurology, Sungkyunkwan University of School of Medicine, Samsung Medical Center, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Republic of Korea. <sup>3</sup>Neuroscience Center, Samsung Medical Center, Seoul, Republic of Korea. <sup>4</sup>Samsung Alzheimer Research Center, Research Institute for Future Medicine, Samsung Medical Center, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Republic of Korea. <sup>5</sup>Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University, Suwon, Republic of Korea. <sup>6</sup>Department of Health Science and Technology, SAHST, Sungkyunkwan University, Seoul, Republic of Korea. <sup>7</sup>Stem Cell and Regenerative Medicine Institute, Samsung Medical Center, Seoul, Republic of Korea. <sup>8</sup>These authors contributed equally: Young Hee Jung and Hyejoo Lee. ✉email: hmjang57@gmail.com; sangwonseo@empal.com

	Total N = 71	A $\beta$ (-) N = 25	A $\beta$ (+) N = 46	P value
<b>Demographics</b>				
Age	74.0 $\pm$ 7.0	72.1 $\pm$ 7.5	75.0 $\pm$ 6.6	0.098
Female	37 (52.1)	15 (60.0)	22 (47.8)	0.327
Education years	9.9 $\pm$ 5.4	8.6 $\pm$ 5.4	10.6 $\pm$ 5.4	0.135
ApoE 4	28 (41.2)	6 (26.1)	22 (48.9)	0.071
ApoE 2	10 (14.7)	2 (8.7)	8 (17.8)	0.317
Hypertension	40 (56.3)	18 (72.0)	22 (47.8)	0.050
Diabetes	15 (21.1)	8 (32.0)	7 (15.2)	0.098
Dyslipidemia	20 (28.2)	10 (40.0)	10 (21.7)	0.102
Cardiac disease	5 (7.0)	0 (0.0)	5 (10.9)	0.087
Previous stroke	16 (22.5)	9 (36.0)	7 (15.2)	0.045
<b>Imaging markers</b>				
Number of lacunes	4.5 $\pm$ 8.8	9.8 $\pm$ 13.1	1.7 $\pm$ 2.5	< 0.001
Presence of cSS	23 (32.4)	3 (12.0)	20 (43.5)	0.007
Presence of lobar ICH	22 (31.0)	11 (44.0)	11 (23.9)	0.080
Number of lobar CMBs	49.5 $\pm$ 69.5	26.3 $\pm$ 33.2	62.2 $\pm$ 80.4	0.037
Number of deep CMBs	2.9 $\pm$ 6.4	4.7 $\pm$ 5.9	1.9 $\pm$ 6.5	0.075
Number of superficial cerebellar CMBs	1.8 $\pm$ 4.7	1.7 $\pm$ 4.4	1.8 $\pm$ 5.0	0.944
Number of CMBs in cerebellar dentate nucleus	0.3 $\pm$ 0.5	0.6 $\pm$ 1.0	0.2 $\pm$ 0.8	0.049

**Table 1.** Clinical characteristics of study participants. \*Numbers are presented mean  $\pm$  standard deviation or n (%). A $\beta$  = amyloid  $\beta$ , CMB = cerebral microbleeds, APO E = apolipoprotein E, cSS = cortical superficial siderosis, ICH = intracerebral hemorrhage.

is also supported by the finding of our previous study that A $\beta$  + CAA patients had a greater burden of CAA MRI markers and a lower burden of hypertensive angiopathy MRI markers such as lacunes<sup>7</sup>. We consider delineating the probable underlying pathology important because it enables better prognostic prediction and appropriate treatment decisions<sup>13,14</sup>. Therefore, predicting A $\beta$  positivity in patients with CAA MRI markers would be clinically useful, because it could help predict prognosis.

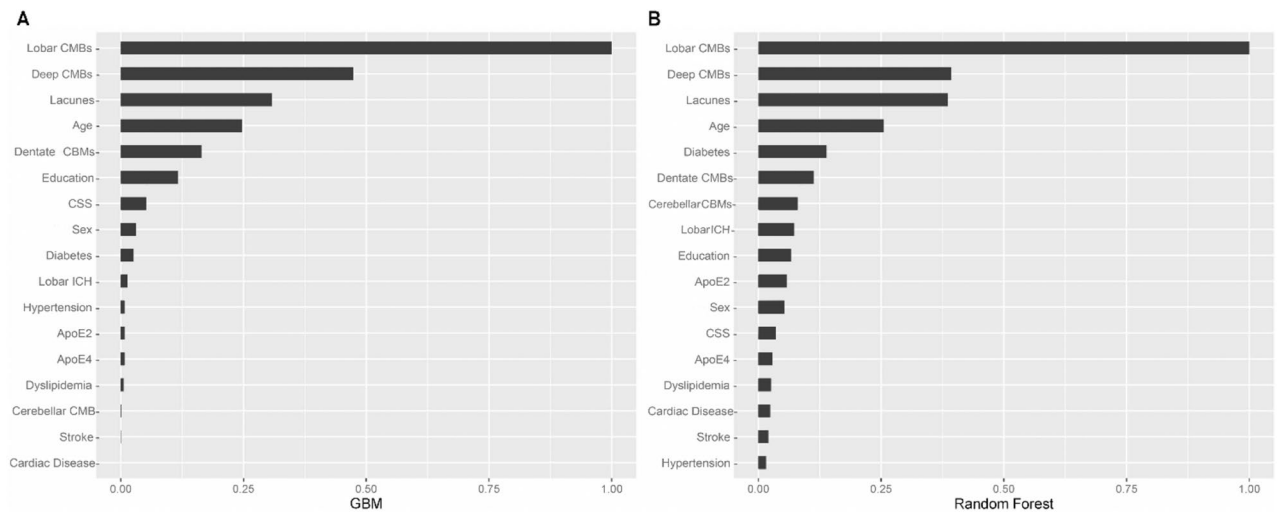
Among prediction models, machine learning methods have been getting much attention due to high predictive power and reliable performance. However, lack of the interpretability of the internal processing has become a major issue in machine learning research. To overcome this limitation, we chose two tree-based machine learning models: gradient boosting machine (GBM)<sup>15</sup> and random forest (RF)<sup>16</sup>. These two methods can effectively quantify the relative importance of variables and provide their cut-off values, which provides clinically meaningful insights.

Therefore, we aimed to identify the most important variables (among imaging markers and clinical characteristics) and the optimal cut-off values of them (such as the number of lobar CMBs) to predict A $\beta$  PET positivity using machine learning based models, in patients with suspected CAA MRI markers. We consider that this prediction model is going to help clinicians to easily select patients with poor prognosis, based on clinical and imaging findings only.

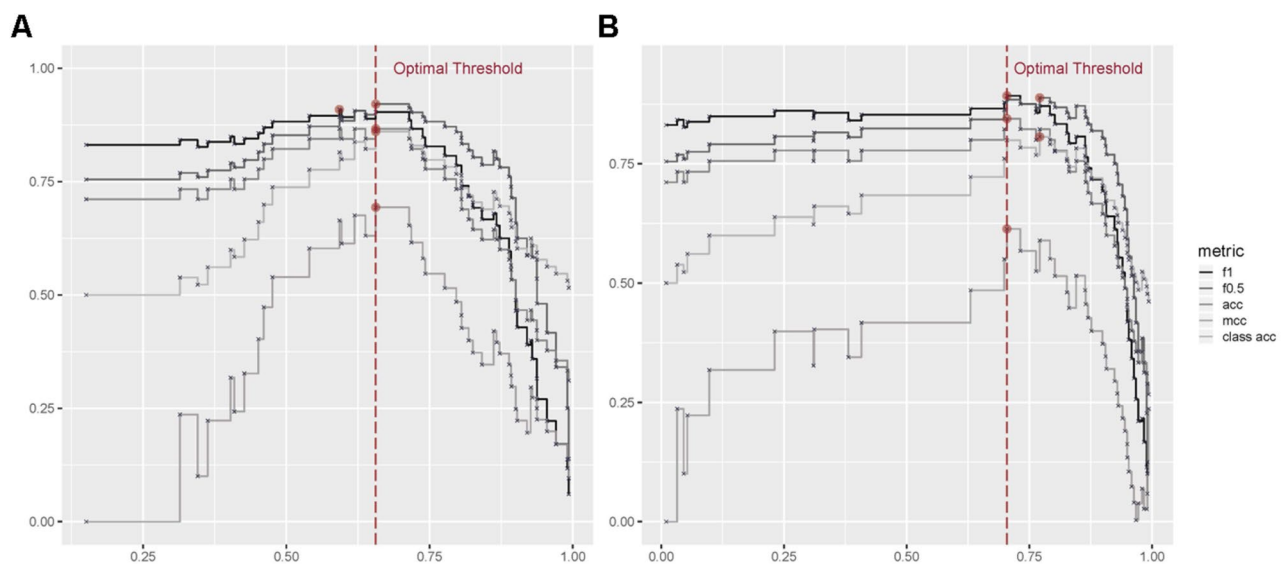
## Results

**Baseline characteristics.** We recruited 71 participants, of whom 25 participants were A $\beta$ - and remaining 46 participants were A $\beta$ +. Mean Age (72.1  $\pm$  7.5 vs. 75.0  $\pm$  6.6,  $p$  = 0.098) and female ratio (15 vs. 22%,  $p$  = 0.327) were not different between the two groups. However, the A $\beta$ - group showed a tendency of a higher prevalence of hypertension (40 vs. 18%,  $p$  = 0.050) and a higher rate of previous stroke (16 vs 9%,  $p$  = 0.045) compared with the A $\beta$  + group. As a surrogate marker of CSVD, the number of lacunes was significantly higher in the A $\beta$ - groups than in the A $\beta$  + group (9.8  $\pm$  13.1 vs 1.7  $\pm$  2.5,  $p$  < 0.001). In terms of CAA markers, cSS was more commonly found in the A $\beta$  + than in the A $\beta$ - group (43.5 vs. 12%,  $p$  = 0.007). Number of lobar CMBs was also higher in the A $\beta$  + group than the A $\beta$ - group (26.3  $\pm$  33.2 vs. 62.2  $\pm$  80.4,  $p$  = 0.037). Although the number of superficial cerebellar CMBs was not different between the A $\beta$ - and A $\beta$  + groups (1.7  $\pm$  4.4 vs. 1.8  $\pm$  5.0,  $p$  = 0.994), the number of CMBs in cerebellar dentate nucleus was higher in the A $\beta$ - group than in the A $\beta$  + group (0.6  $\pm$  1.0 vs. 0.2  $\pm$  0.8,  $p$  = 0.049) (Table 1).

**Important predictive variables for A $\beta$  positivity.** Among 17 clinical and imaging variables, we computed relative importance using GBM and RF algorithms and selected the most important variables, which were similar in both models. The five important variables ranked in GBM model and their relative importance are as follows: the number of lobar CMBs (18.6), the number of deep CMBs (8.8), the number of lacunes (5.7), age (4.6), and the number of CMBs in dentate nucleus (3.1). On the other hand, RF model chose the six important variables as follows: the number of lobar CMBs (60.4), the number of deep CMBs (23.7), the number of lacunes (23.3), age (15.4), the absence of diabetes (8.4), and the number of CMBs in the dentate nucleus (6.8) (Fig. 1).



**Figure 1.** Importance plot of variables in GBM and RF models. GBM = gradient boosting model, RF = random forest, CMB = cerebral microbleed, cSS = cortical superficial siderosis, ICH = intracerebral hemorrhage, HTN = hypertension, APOE = apolipoprotein E. (A) Among 17 variables which were associated with CAA, the five important variables ranked in GBM model (relative importance) are as follows: the number of lobar CMBs (18.6), the number of deep CMBs (8.8), the number of lacunes (5.7), age (4.6), the number of CMBs in dentate nucleus. (3.1) (B) RF model chose the six important variables (relative importance) are as follows: the number of lobar CMBs (60.4), the number of deep CMBs (23.7), the number of lacunes (23.3), age (15.4), The absence of diabetes (8.4), and the number of CMBs in dentate nucleus (6.8).



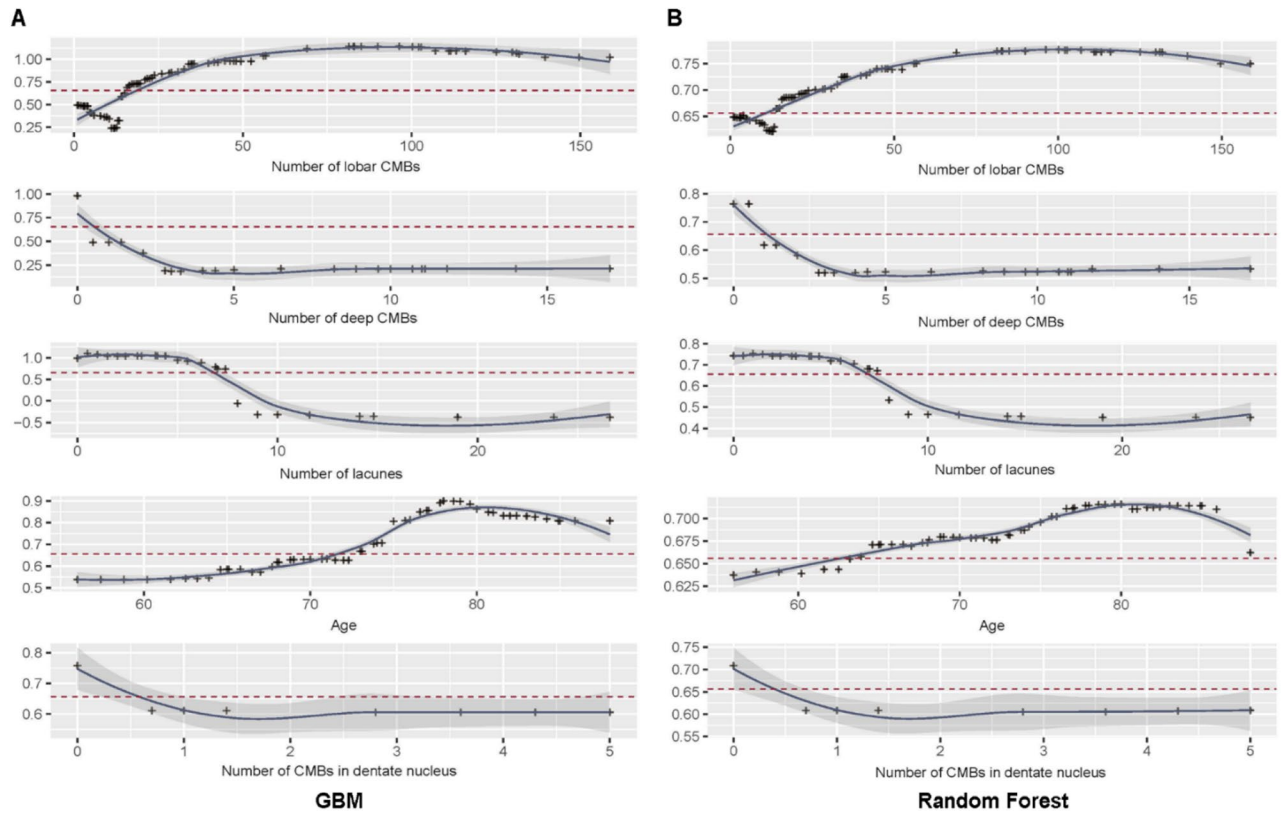
**Figure 2.** Detecting optimal threshold in multiple change points in GBM and RF models. GBM = gradient boosting model, ACC = accuracy, MCC = misclassification, Class ACC = class per accuracy, F1 = harmonic mean of the positive and negative predictive values with equal weights, F0.5 = mean of positive and negative predictive values, which gives more weight to PPV than to NPV. (A) In GBM method, the threshold was determined as 0.7043, when four metrics (F0.5, ACC, MCC, class ACC) are at their maximum values, respectively. (B) In RF method, threshold was determined as 0.6561, when three metrics (f1, ACC, MCC) are at their maximum values, respectively.

After adding the lobar CMB/deep CMB ratio as a new variable, the highly ranked variables and their performance remained almost the same as the original result. (Supplementary Table S2).

**Cut-off values of predictive variables for A $\beta$  positivity.** In GBM, the threshold was determined as 0.7043 when four metrics (F0.5, ACC, MCC, class ACC) were at their maximum values respectively. In RF, threshold was determined as 0.6561, when three metrics (F1, accuracy, misclassification) were at their maximum values, respectively. (Fig. 2) Using these thresholds obtained as above, we determined cut-off values of important

	GBM	Random forest
Number of lobar CMBs	16.4	14.0
Number of deep CMBs	0	0
Number of lacunes	7.4	7.4
Age	74.3	63.9
Number of CMBs in dentate nucleus	0	0

**Table 2.** Cut-off values of GBM and RF models. GBM = gradient boosting model, CMB = cerebral microbleed.



**Figure 3.** Cut-off values of important variables in GBM and RF models. GBM = gradient boosting machines, RF = random forest, CMBs = cerebral microbleeds, In PDP curve (y is threshold of metrics, and x is cut-off value), the optimal cut-off value was determined, when the curve passes the threshold which was obtained above. **(A)** Cut-off values of variables to predict A $\beta$  positivity in GBM were as follows: (1) If the number of lobar CMB is more than 16.4, (2) if there is no deep CMBs, (3) if the number of lacunes is more than 7.4, (4) if age is older than 74.3 (GBM), (5) if there is no CMBs in dentate nucleus. **(B)** Cut-off values of variables to predict A $\beta$  positivity in RF were as follows: (1) If the number of lobar CMB is more than 14.3, (2) if there is no deep CMBs, (3) if the number of lacunes is more than 7.4, (4) if age is older than 64, (5) if there is no CMBs in dentate nucleus.

variables. Cut-off values of variables to predict A $\beta$  positivity were as follows: (1) If the number of lobar CMB is more than 16.4 (GBM)/14.3 (RF), (2) If there is no deep CMBs (GBM and RF), (3) If the number of lacunes is more than 7.4 (GBM and RF), (4) If age is older than 74.3 (GBM) /64(RF), (5) If there is no CMBs in dentate nucleus (Table 2, Fig. 3).

**Model performances of GBM and RF based prediction models.** Both GBM and RF models showed good performances; MSE was  $0.14 \pm 0.02$  in GBM and  $0.18 \pm 0.06$  in RF. RMSE was  $0.41 \pm 0.08$  in GBM and  $0.37 \pm 0.03$  in RF. Logarithmic loss was  $0.47 \pm 0.07$  in GBM and  $0.53 \pm 0.17$  in RF. Mean per class error was  $0.22 \pm 0.06$  in GBM and  $0.25 \pm 0.14$  in RF. Gini impurity was  $0.65 \pm 0.09$  in GBM and  $0.60 \pm 0.24$  in RF. AUC was  $0.83 \pm 0.04$  in GBM and  $0.80 \pm 0.12$  in RF. Precision-recall AUC was  $0.86 \pm 0.04$  in GBM and  $0.67 \pm 0.18$  in RF. (Table 3).

Performance measures (mean $\pm$ SD <sup>a</sup> )	GBM	Random forest
Mean square error	0.14 $\pm$ 0.02	0.18 $\pm$ 0.06
Root mean square error	0.37 $\pm$ 0.03	0.41 $\pm$ 0.08
Logarithmic loss	0.47 $\pm$ 0.07	0.53 $\pm$ 0.17
Mean per class error	0.22 $\pm$ 0.06	0.25 $\pm$ 0.14
Gini impurity	0.65 $\pm$ 0.09	0.60 $\pm$ 0.24
Area under curve	0.83 $\pm$ 0.04	0.80 $\pm$ 0.12
Precision-recall area under curve	0.86 $\pm$ 0.04	0.67 $\pm$ 0.18

**Table 3.** The performance of GBM and RF models. <sup>a</sup>Lower values of mean square error, root mean square error, logarithmic loss, mean per class error, and gini impurity means better prediction power, and lower SD means higher reliability. GBM = gradient boosting model; SD = standard deviation.

## Discussion

In present study, we developed machine-learning based models to predict A $\beta$  positivity on PET in patients with suspected CAA markers. Our first major finding was that GBM and RF algorithms consistently ranked anatomical distribution of CMBs, age, the number of lacunes as the most important variables for predicting A $\beta$  PET positivity. Our second major finding was the suggested cut-off values of these important variables (particularly, lobar CMBs higher than at least 14 and lacune number less than 7.4) predicting A $\beta$  PET positivity. Finally, both models showed good performances, but GBM-based model performance was slightly better than RF-based model.

The first major finding was that both machine learning methods consistently ranked the number of lobar CMBs, deep CMBs, lacunes, and dentate nucleus CMBs, and age as the most important variables for prediction of A $\beta$  PET positivity. Besides, the ranked orders of variables were similar in two models, although the value of relative importance was slightly different. Among selected variables, topographic location of CMBs (high number of lobar CMBs and absence of deep CMBs) and old age are well-known CAA predicting features according to the modified Boston Criteria. Therefore, it is reasonable that these factors could also predict A $\beta$  PET positivity. However, the presence of cSS, which is considered as one of the important imaging parameters of CAA, was not highly ranked in our models. Considering that the prevalence of cSS was significantly higher in the A $\beta$  + group than A $\beta$ - group in our study, cSS seems to be associated with A $\beta$  positivity, which is also consistent with previous study<sup>17</sup>. Nevertheless, cSS could be a less important predicative variable than topographic distribution of CMBs, number of lacunes, and age, when the model is made by the combination of various features in the memory clinic, which might be attributed to small number of patients having cSS in this clinical circumstance.

The second major findings were the cut-off values of important variables to predict A $\beta$  positivity. First, both machine learning-based models showed that no CMB in deep structures and cerebellar dentate nucleus was predictive of A $\beta$  positivity. We consider that our study finding supports the modified Boston criteria in which the presence of deep CMBs is exclusion criteria for probable CAA, even when number of lobar CMBs outweighs that of deep CMBs as in our cases; If these cases were advanced CAA as we hypothesized, presence of deep CMBs might not lower the possibility of A $\beta$  positivity. This is along the same line with cerebellar dentate nucleus involvement. As equivalents of deep CMBs, CMBs in cerebellar dentate nucleus are likely due to hypertensive angiopathy as reported in recent studies<sup>18,19</sup>. Our prediction models suggested optimal cut-off values of lobar CMBs predicting A $\beta$  positivity as 16.4 (GBM) or 14.3 (RF). Although, the modified Boston Criteria proposed that at least 2 lobar hemorrhages were enough to be diagnosed with probable CAA, some might argue that only two lobar CMBs could be found incidentally without CAA pathology. In addition, the previous study including patients with only CMBs reported that higher CMB counts increased specificity for predicting CAA<sup>20</sup>. This suggests the additional possibility that likelihood of CAA increases in a proportional relationship with CMBs number rather than a sharp threshold at  $\geq 2$  CMBs<sup>20</sup>. From this perspective, the cut-off of lobar CMBs (particularly in the absence of symptomatic lobar ICHs) for predicting A $\beta$  positivity must be higher than two considering that A $\beta$  positivity may relate to advanced CAA pathology. Therefore, machine learning methods derived cut-off values of lobar CMBs could be usefully applied to predict A $\beta$  positivity, which is associated with CAA pathology and poor clinical prognosis in patients with only multiple CMBs even in the absence of symptomatic lobar ICHs.

Other noteworthy findings were the cut-off values of age and lacunes; The age cut-off for predict A $\beta$  positivity were 74.3 (GBM) and 63.9 (RF), which were older than 55 years as presented in the modified Boston criteria. Although minimum age at CAA could develop is 55 years old according to the suggested criteria, this result shows that an older age increases the possibility of A $\beta$  positivity in patients with CAA MRI markers. Finally, both prediction models showed that the number of lacunes lower than 7.4 was predictive of A $\beta$  positivity. We consider that lacunes were considered as surrogate marker of hypertensive angiopathy rather than CAA. Therefore, when patients have mixed deep and lobar CMBs, the number of lacunes higher than 7.4 is almost always suggestive of hypertensive angiopathy, which is more likely to have negative A $\beta$  PET scans. Nevertheless, the cut-off value of 7.4 was higher than expectation, which we considered was because FLAIR image with axial thickness of 2 mm (which is fivefold thinner than usual thickness of 10 mm) enabled sensitive counting of lacunes in study patients.

The final major finding was that both machine learning-based models showed good performance with higher than 80% of predictive accuracy, although GBM was slightly better than RF. We selected GBM and RF for the following reasons. First, previous large-scale studies have consistently suggested GBM and RF as robust ML algorithms<sup>21–23</sup>. Second, the generalizability may be ensured by comparing two methods with complementary methodological backgrounds. For example, although GBM performed better than RF on the skewed data, it



could provide misleading outcomes from the noisy data and vice versa. Third, for more reliable predictions, we selected tree-based ML models and compared their interpretable predictions. Tree-based models provide the same interpretable methods such as relative importance and PDP. Variable importance determines the features that influence accurate classification<sup>24</sup>. Besides, PDP can estimate whether the variables had a positive or negative effect on the prediction using a marginal distribution. Thus, the intersection between the negative and positive PDPs provides cut-off values of the variables. In this study, GBM and RF showed similar interpretable results.

Especially, the importance of topographic distribution was reconfirmed by our machine learning methods. Particularly, new cut-off values of lobar CMBs and age in present study could be used as a supportive measure to predict A $\beta$  positivity in patients with CAA MRI markers. Furthermore, the diagnosis of patients with many lobar CMBs combined with a few deep CMBs has been unclear. However, these models enable us to distinguish A $\beta$  pathology from hypertensive angiopathy in this population by predicting A $\beta$  positivity. We can also predict A $\beta$  positivity using clinical information and MR imaging, which is less expensive and more readily available. Methodologically, the cut-off values have conventionally obtained, using receiver operating characteristic curve with only two metrics, sensitivity, and specificity. However, in this study, we obtained cut-off values using five metrics, which enabled higher dimensional analyses and consequently better accuracy than the conventional approach.

Our study has its strength in two machine learning based models (GBM and RF) which showed consistent and reliable results with good performances, although they independently select important variables and rank the important variables in supervised ways. We acknowledge some limitations of this study. There may exist concern about the overfitting problem in training models with a relatively small number of data samples. In addition, the cutoff values for variables such as age could have been biased due to the small sample size although we considered that the demographic data and the imaging features of this study population were reflective of the characteristics of patients with CAA markers who visited memory clinics. Therefore, future studies are required to develop more generalizable models with a possible external dataset. Also, we used A $\beta$  positivity on PET instead of a pathologic confirmation. Nevertheless, prediction for A $\beta$  positivity would be useful for clinicians to understand their clinical courses, based on clinical significance of A $\beta$  PET positivity in CAA patients<sup>7</sup>. Finally, we used three different A $\beta$  PET ligands in this study. However, this limitation may have been overcome, as previous studies demonstrated that three different PET uptakes are highly correlated with each other<sup>25–27</sup>.

In conclusion, we developed two reliable machine learning-based models to predict A $\beta$  positivity in 71 patients with suspected CAA MRI markers using various clinical and imaging features, and they suggested useful clinical cut-offs for predictive variables. These models may help clinician to predict prognosis of patients with suspected CAA markers and to make stratified enrollment in clinical trials, by predicting A $\beta$  PET positivity.

## Methods

**Participants.** We included all 2333 patients who visited our memory clinic (Samsung Medical Center, Korea), complaining of cognitive impairment and underwent A $\beta$  PET from September 2008 to June 2018. We scrutinized Brain MRI of all patients, and recruited patients who met the following criteria that we developed in this study: (1) If patients have at least one lobar ICH or cSS, only one lobar CMB is enough for them to be included; (2) If patients do not have either lobar ICH or cSS, 10 or more lobar CMBs are required for them to be included; (3) If patients have both lobar and deep CMBs, the number of lobar CMBs should be higher than that of deep CMBs. Therefore, we finally included 71 patients (26 PiB PET, 43 florbetaben PET, 2 flutemetamol PET) whom we refer to as “patients with suspected CAA markers” in this study.

We excluded patients with the presence of secondary causes of cognitive deficit (e.g. vitamin B12/folate, syphilis serology, and/or thyroid dysfunction), or structural lesion except for lobar ICH (e.g. territorial cerebral infarctions and brain tumors), or with psychiatric illness such as schizophrenia.

The Institutional Review Board of Samsung Medical Center approved the study protocol and informed consent was obtained from all subjects or, if subjects are under 18, from a parent and/or legal guardian. This manuscript does not contain information or image that can lead to identification of a study participant. The methods were carried out in accordance with the approved guidelines.

**MR image acquisition.** All participants underwent brain MRI including T2\* GRE and fluid attenuated inversion recovery (FLAIR). The following parameters were used for the T2\* GRE images: axial slice thickness, 5.0 mm; inter-slice thickness, 2 mm; repetition time (TR), 669 ms; echo time (TE) 16 ms; flip angle, 18°; matrix size, 560 × 560 pixels. The following parameters were used for the 3D FLAIR images: axial slice thickness of 2 mm; no gap; repetition time of 11 000 ms; echo time of 125 ms; flip angle of 90°; and matrix size of 512 × 512 pixels.

**Assessment of CMB, cSS, lobar ICH and lacunes on MRI.** Imaging analysis was carried out by individuals who were trained in neuroimaging rating and blinded to the participant clinical details. All structural imaging markers of CSVD were rated in accordance with consensus guidelines<sup>28,29</sup>. Lobar CMBs were defined as homogenous and round lesions with signal loss ( $\leq 10$  mm in diameter) on T2\* GRE images, with location in exclusively lobar (cortex, gray-white matter junction, subcortical white matter) areas. Deep CMBs were defined as CMB in basal ganglia gray matter, internal and external capsules, and thalamus, according to brain observer microbleed scale (BOMBS)<sup>30</sup>. Infratentorial CMBs were also classified as deep CMBs. Cerebellar CMBs were separately counted and classified into dentate nucleus and superficial cerebellar CMBs<sup>18</sup>. cSS was defined as linear hypointensities on T2\* GRE images consistent with chronic blood residues in the superficial layers of the cerebral cortex<sup>31</sup>. Lacunes were identified and counted in accordance with STRIVE (STandards for Reporting Vascular changes on nEuroimaging)<sup>28</sup>.

**A $\beta$  PET imaging acquisition.** The mean value of MRI-A $\beta$  PET interval was  $8.8 \pm 9.8$  months. All patients underwent A $\beta$  PET using a Discovery STe PET/CT scanner (GE Medical Systems, Milwaukee, WI) in a 3D scanning mode that examined 47 slices of 3.3 mm thickness spanning the entire brain. A 16-slice helical CT (140 keV, 80 mA; 3.75 mm section width) was performed for attenuation correction. For 11C-PiB PET, a 30-min emission static PET scan was performed 60 min after injection into an antecubital vein as a bolus of a mean dose of 420 MBq. For 18F-Florbetaben PET, a 20-min emission PET scan with dynamic mode (consisting of  $4 \times 5$  min frames) was performed 90 min after injection into an antecubital vein as a bolus of a mean dose of 381 MBq. For flutemetamol, 20-min emission static PET scan with dynamic mode (consisting of  $4 \times 5$  min frames) was performed 90-min after injection into an antecubital vein as a bolus of a mean dose of 185 MBq.

**A $\beta$  PET image preprocessing and definition for A $\beta$  positivity.** Both MR and A $\beta$  PET images were co-registered with each other using the rigid-body transformation. The T1-weighted MR image of each subject was aligned with the MNI-152 template using a non-linear deformation including translation, rotation, scaling and shearing. After standard space registration, we divided grey matter into 116 regions using the Automated Anatomical Labeling atlas<sup>32</sup>. In order to compute standardized uptake value ratios (SUVR) for PiB and florbetaben<sup>33</sup>, every voxel intensity was normalized by the mean intensity of cerebellum regions. For flutemetamol PET, we computed SUVR by the mean intensity of pons regions as reference value. We defined A $\beta$  positivity on each PET as follows: (1) If global PiB SUVR (assessed from the volume-weighted average SUVR of 28 bilateral cerebral cortical VOIs) was greater than 1.5, (2) If visual rating score on florbetaben PET was 2 or 3 on the brain A $\beta$  plaque load (BAPL) scoring system<sup>34</sup>, or (3) If any one of the brain regions systematically reviewed for <sup>18</sup>F-flutemetamol PET was positive in either hemisphere<sup>34</sup>.

**Statistical analysis.** We compared the demographic and clinical characteristics between the A $\beta$ + and A $\beta$ - groups using Student t-tests for continuous variables and chi-square test for dichotomous variables. Statistical analyses were performed using R version 3.5.0.

**Potential variables for predicting A $\beta$  positivity.** We included all clinical and imaging characteristics as potential variables in model development: gender, education year, vascular risk factors (dyslipidemia, diabetes, cardiac disease, previous stroke, and hypertension), apolipoprotein E (APOE) genotype, number of CMBs in each location (number of lobar CMBs, deep CMBs, dentate nucleus CMBs, and superficial cerebellar CMBs), presence of lobar ICH and cSS, and number of lacunes.

For sensitivity analysis, we performed the same analysis with the lobar CMB/deep CMB ratio as an additional variable. If the number of deep CMBs was zero, we used the number of lobar CMBs instead of the lobar CMB/deep CMB ratio.

**Model generation for classifying A $\beta$  positivity.** Among the tree-based ML models, we selected GBM and RF. GBM generates accurate classifiers using linear combinations of the base classifiers adjusted by their weights iteratively. The PDP approach was originally introduced by J.H. Friedman in the GBM paper<sup>15</sup>. RF creates multiple decision trees using bootstrap samples and the binning of outliers. RF aggregates their decisions by averaging or majority voting<sup>35</sup>. GBM and RF analysis were carried out using different combinations of hyperparameter settings and varying search criteria in randomly selected trials. In grid search process, the advanced computing power enabled searching the entire hyperparameter space.

Twenty repetitions of tenfold cross-validation (CV) were conducted in order to select the optimal solution<sup>36</sup>. K-fold CV is to divide the data set into non-overlapping k equal partitions. Each data partition is then used as the validation set and the remaining K-1 partitions are used as a training set. We selected K = 10 as an empirically ideal situation of 10 training sets and 10 validation sets<sup>37</sup>. Under the CV procedure, the generalization of predictive power and validation errors were computed. The best parameter setting corresponding to the minimal error obtained by CV, was then applied to train the model using a train set and a validation set, which were 70% and 15% of the entire data set respectively. Remaining 15% data set was used as a test set, and their performance was estimated. The whole process was repeated over 20 times in order to evaluate reliable classifier performances.

Although we selected GBM and RF in this study, we compared the performance of other ML methods such as logistic regression<sup>38</sup>, k-nearest neighbors (KNN)<sup>39</sup>, and support vector machine (SVM)<sup>40</sup>. Further details on the classifiers are provided in Supplementary Method 1. For a fair comparison, the same CV data partitions were used across all the ML models, and performance was estimated using the arithmetic means of the outcome. Supplementary Table S1 and Supplementary Method 2 provide the details on model performance and additional performance measures, respectively.

**Interpretable machine learning.** For each analysis, the extent to which the variables influenced the accuracy of classification was quantified by calculating the relative variable importance<sup>41</sup>. In the tree-based model such as GBM and RF, when the variable split the tree, relative importance value of that variable was estimated by discrepancy of the squared error loss over all tree. A higher relative importance value indicates greater influence of the variable in classifying A $\beta$  positivity.

Optimal threshold was estimated in terms of F1 score, F0.5 score, accuracy, misclassification and class accuracy which are widely recommended for classification tasks<sup>42</sup>.

The cut-off values of the important, numerical variables were determined by partial dependence plot (PDP) which is a graphical representation tool describing the relationship between target feature and input features

resulting by importance variables. Let  $\mathbf{x}$  be the space of input variables consisting of a chosen subset space and its complementary space,

$$x_s \cup x_c = \mathbf{x}$$

Then the approximation  $\hat{F}(\mathbf{x})$  depend on both subset space.

$$\hat{F}(\mathbf{x}) = \hat{F}(x_s, x_c), \hat{F}_c(x_s) = \hat{F}(x_s | x_c)$$

$$\bar{F}_s(x_s) = E_{x_c}[\hat{F}(\mathbf{x})] = \int \hat{F}(x_s, x_c) p_c(x_c) dx_c$$

$$\tilde{F}_s(x_s) = E_x[\hat{F}(\mathbf{x}) | x_s] = \int \hat{F}(\mathbf{x}) p_z(x_c | x_s) dx_c$$

In PDP curve ( $y$  is threshold of metrics, and  $x$  is cut-off value), the optimal cut-off value was determined, when the curve passes the threshold which was obtained above.

**Assessment of model performance.** To assess model performance of prediction model, we used six measures as follows: mean square error (MSE), root mean square error (RMSE), logarithmic loss, mean per class error, area under curve, precision-recall area under curve (AUC), gini impurity. We computed the mean values of each measure after 20 iterations.

The MSE of estimator (of a procedure for estimating an unobserved quantity) measures the average of the square of the error- that is, the average squared difference between the estimated values and the actual value. The less MSE means better prediction. The tracking task was scored by calculating the RMSE between the target and response signals<sup>43</sup>. Logarithmic loss (related to cross-entropy) increases as the predicted probability diverges from the actual label. Mean Per Class Error is the average of the errors of each class in multi-class data set, which measures misclassification of the data across the classes. AUC is used to evaluate how well a binary classification model can distinguish true positives from false positives. Precision-Recall curves summarize the trade-off between the true positive rate and the positive predictive value for a predictive model using different probability thresholds especially for imbalanced dataset. Gini impurity is a measure of how often a randomly chosen element from the set would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset.

## Data availability

The data sets generated or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 3 April 2020; Accepted: 13 October 2020

Published online: 02 November 2020

## References

- Exley, C. & Esiri, M. Severe cerebral congophilic angiopathy coincident with increased brain aluminium in a resident of Camelford, Cornwall, UK. *J. Neurol. Neurosurg. Psychiatry* **77**, 877–879 (2006).
- Biffi, A. & Greenberg, S. M. Cerebral amyloid angiopathy: A systematic review. *J. Clin. Neurol.* **7**, 1–9 (2011).
- Knudsen, K. A., Rosand, J., Karluk, D. & Greenberg, S. M. Clinical diagnosis of cerebral amyloid angiopathy: Validation of the Boston criteria. *Neurology* **56**, 537–539 (2001).
- Linn, J. *et al.* Prevalence of superficial siderosis in patients with cerebral amyloid angiopathy (CME). *Neurology* **74**, 1346–1350. <https://doi.org/10.1212/WNL.0b013e3181dad605> (2010).
- Raposo, N. & Sonnen, J. A. Amyloid-PET in cerebral amyloid angiopathy: Detecting vascular amyloid deposits, not just blood. *Neurology* <https://doi.org/10.1212/wnl.0000000000004548> (2017).
- Farid, K., Charidimou, A. & Baron, J. C. Amyloid positron emission tomography in sporadic cerebral amyloid angiopathy: A systematic critical update. *NeuroImage Clin.* **15**, 247–263. <https://doi.org/10.1016/j.nicl.2017.05.002> (2017).
- Jang, H. *et al.* Clinical significance of amyloid beta positivity in patients with probable cerebral amyloid angiopathy markers. *Eur. J. Nucl. Med. Mol. Imaging* **46**, 1287–1298. <https://doi.org/10.1007/s00259-019-04314-7> (2019).
- Okello, A. *et al.* Conversion of amyloid positive and negative MCI to AD over 3 years An 11C-PIB PET study. *Neurology* **73**, 754–760 (2009).
- Park, J. H. *et al.* Effects of cerebrovascular disease and amyloid beta burden on cognition in subjects with subcortical vascular cognitive impairment. *Neurobiol. Aging* **35**, 254–260. <https://doi.org/10.1016/j.neurobiolaging.2013.06.026> (2014).
- Kim, H. J. *et al.* Relative impact of amyloid-beta, lacunes, and downstream imaging markers on cognitive trajectories. *Brain* **139**, 2516–2527. <https://doi.org/10.1093/brain/aww148> (2016).
- Kim, H. J. *et al.* Clinical effect of white matter network disruption related to amyloid and small vessel disease. *Neurology* **85**, 63–70. <https://doi.org/10.1212/WNL.0000000000001705> (2015).
- Thal, D. R., Ghebremedhin, E., Orantes, M. & Wiestler, O. D. Vascular pathology in Alzheimer disease: Correlation of cerebral amyloid angiopathy and arteriosclerosis/lipohyalinosis with cognitive decline. *J. Neuropathol. Exp. Neurol.* **62**, 1287–1301 (2003).
- Thon, J. M. & Gurol, M. E. Intracranial hemorrhage risk in the Era of antithrombotic therapies for ischemic stroke. *Curr. Treat. Options Cardiovasc. Med.* **18**, 29. <https://doi.org/10.1007/s11936-016-0453-y> (2016).
- Lauer, A., Greenberg, S. M. & Gurol, M. E. Statins in intracerebral hemorrhage. *Curr. Atheroscler Rep.* **17**, 46. <https://doi.org/10.1007/s11883-015-0526-5> (2015).
- Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
- Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
- Na, H. K. *et al.* Cortical superficial siderosis: A marker of vascular amyloid in patients with cognitive impairment. *Neurology* **84**, 849–855. <https://doi.org/10.1212/wnl.0000000000001288> (2015).



18. Renard, D., Tatu, L. & Thouvenot, E. Infratentorial cerebral microbleeds in patients with cerebral amyloid angiopathy. *J. Stroke Cerebrovasc. Dis.* **27**, 2534–2537. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2018.05.015> (2018).
19. Pasi, M. *et al.* Cerebellar microbleed distribution patterns and cerebral amyloid angiopathy. *Stroke* **50**, 1727–1733. <https://doi.org/10.1161/strokeaha.119.024843> (2019).
20. Martinez-Ramirez, S. *et al.* Diagnostic value of lobar microbleeds in individuals without intracerebral hemorrhage. *Alzheimer's Dementia* **11**, 1480–1488. <https://doi.org/10.1016/j.jalz.2015.04.009> (2015).
21. Caruana, R. & Niculescu-Mizil, A. In *Proceedings of the 23rd International Conference on Machine Learning*. 161–168 (ACM).
22. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. (Springer, Berlin, 2009).
23. Chu, A. *et al.* A decision support system to facilitate management of patients with acute gastrointestinal bleeding. *Artif. Intell. Med.* **42**, 247–259 (2008).
24. Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. Bayesian measures of model complexity and fit. *J. R. Statist. Soc. B* **64**, 583–639. <https://doi.org/10.1111/1467-9868.00353> (2002).
25. Landau, S. *et al.* Amyloid PET imaging in Alzheimer's disease: A comparison of three radiotracers. *Eur. J. Nuclear Med. Mol. Imaging* **41**, 1398–1407 (2014).
26. Cho, S. H. *et al.* Head-to-head comparison of 18F-florbetaben and 18F-flutemetamol in the cortical and striatal regions. *J. Alzheimer's Dis.*, **76**, 281–290 (2020).
27. Cho, S. H. *et al.* Concordance in detecting amyloid positivity between 18F-florbetaben and 18F-flutemetamol amyloid PET using quantitative and qualitative assessments. <https://doi.org/10.21203/rs.3.rs-29049/v1> (2020).
28. Wardlaw, J. M. *et al.* Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol.* **12**, 822–838. [https://doi.org/10.1016/S1474-4422\(13\)70124-8](https://doi.org/10.1016/S1474-4422(13)70124-8) (2013).
29. Greenberg, S. M. *et al.* Cerebral microbleeds: A guide to detection and interpretation. *Lancet Neurol.* **8**, 165–174 (2009).
30. Cordonnier, C. *et al.* Improving interrater agreement about brain microbleeds: Development of the Brain Observer MicroBleed Scale (BOMBS). *Stroke* **40**, 94–99. <https://doi.org/10.1161/strokeaha.108.526996> (2009).
31. Linn, J. *et al.* Subarachnoid hemosiderosis and superficial cortical hemosiderosis in cerebral amyloid angiopathy. *AJNR Am. J. Neuroradiol.* **29**, 184–186. <https://doi.org/10.3174/ajnr.A0783> (2008).
32. Tzourio-Mazoyer, N. *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273–289. <https://doi.org/10.1006/nimg.2001.0978> (2002).
33. Lyoo, C. H. *et al.* Cerebellum can serve as a pseudo-reference region in Alzheimer disease to detect neuroinflammation measured with PET radioligand binding to translocator protein. *J. Nucl. Med.* **56**, 701–706. <https://doi.org/10.2967/jnumed.114.146027> (2015).
34. Barthel, H. *et al.* Cerebral amyloid-beta PET with florbetaben (18F) in patients with Alzheimer's disease and healthy controls: A multicentre phase 2 diagnostic study. *Lancet* **10**, 424–435. [https://doi.org/10.1016/S1474-4422\(11\)70077-1](https://doi.org/10.1016/S1474-4422(11)70077-1) (2011).
35. Breiman, L. Bagging predictors. *Mach. Learn.* **24**, 123–140 (1996).
36. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Intern. Joint Conf. of Artificial Intelligence* **2**, 1137–1145 (1995).
37. Bengio, Y. & Grandvalet, Y. No unbiased estimator of the variance of k-fold cross-validation. *J. Mach. Learn. Res.* **5**, 1089–1105 (2004).
38. Cox, D. R. The regression analysis of binary sequences. *J. R. Stat. Soc.: Ser. B (Methodol.)* **20**, 215–232 (1958).
39. Altman, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **46**, 175–185 (1992).
40. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
41. Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc.: Ser. B* **64**, 583–639 (2002).
42. Candel, A., Parmar, V., LeDell, E. & Arora, A. Deep learning with H2O, 5–57 (2016).
43. Hogan, H. & Szeto, A. A simulator for objectively evaluating prospective drivers of the Scott van. *Bull. Prosthet. Res.* **10**, 19–27 (1982).

## Acknowledgements

This research was supported by a fund (2018-ER6203-02) by research of Korea Centers for Disease Control and Prevention, the Brain Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (NRF-2018M3C7A1056512) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2019R1A5A2027340) and (NRF-2020R1A2C1009778).

## Author contributions

Conception and design of the study: H. L., Y. H. J., H. J. and S. W. S. Acquisition and analysis data: H. L., Y. H. J., H. J., H. J. K., D. L. N., and S. W. S. Drafting and revising manuscript for intellectual contents: H. L., Y. H. J., H. J., H. J. K., D. L. N., H. J. H. and S. W. S.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-75664-8>.

**Correspondence** and requests for materials should be addressed to H. J. or S. W. S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020