



Original Article

Easily created prediction model using deep learning software (Prediction One, Sony Network Communications Inc.) for subarachnoid hemorrhage outcomes from small dataset at admission

Masahito Katsuki, Yukinari Kakizawa, Akihiro Nishikawa, Yasunaga Yamamoto, Toshiya Uchiyama

Department of Neurosurgery, Suwa Red Cross Hospital, Suwa, Nagano, Japan.

E-mail: Masahito Katsuki - ktk1122nigt@gmail.com; *Yukinari Kakizawa - ykakisawajp@yahoo.co.jp; Akihiro Nishikawa - aki.west@gmail.com; Yasunaga Yamamoto - yamamotoyasunaga@gmail.com; Toshiya Uchiyama - u_tosh@gmail.com



*Corresponding author:

Yukinari Kakizawa,
Department of Neurosurgery,
Suwa Red Cross Hospital, Suwa,
Nagano, Japan.

ykakisawajp@yahoo.co.jp

Received : 13 September 2020

Accepted : 15 October 2020

Published : 06 November 2020

DOI

10.25259/SNI_636_2020

Quick Response Code:



ABSTRACT

Background: Reliable prediction models of subarachnoid hemorrhage (SAH) outcomes are needed for decision-making of the treatment. SAFIRE score using only four variables is a good prediction scoring system. However, making such prediction models needs a large number of samples and time-consuming statistical analysis. Deep learning (DL), one of the artificial intelligence, is attractive, but there were no reports on prediction models for SAH outcomes using DL. We herein made a prediction model using DL software, Prediction One (Sony Network Communications Inc., Tokyo, Japan) and compared it to SAFIRE score.

Methods: We used 153 consecutive aneurysmal SAH patients data in our hospital between 2012 and 2019. Modified Rankin Scale (mRS) 0–3 at 6 months was defined as a favorable outcome. We randomly divided them into 102 patients training dataset and 51 patients external validation dataset. Prediction one made the prediction model using the training dataset with internal cross-validation. We used both the created model and SAFIRE score to predict the outcomes using the external validation set. The areas under the curve (AUCs) were compared.

Results: The model made by Prediction One using 28 variables had AUC of 0.848, and its AUC for the validation dataset was 0.953 (95%CI 0.900–1.000). AUCs calculated using SAFIRE score were 0.875 for the training dataset and 0.960 for the validation dataset, respectively.

Conclusion: We easily and quickly made prediction models using Prediction One, even with a small single-center dataset. The accuracy of the model was not so inferior to those of previous statistically calculated prediction models.

Keywords: Artificial intelligence, Deep learning, Machine learning, Prediction model, Subarachnoid hemorrhage

INTRODUCTION

Subarachnoid hemorrhage (SAH) due to the ruptured cerebral aneurysm is one of the severe stroke types, and it has a mortality rate of up to 35%. Furthermore, about one-third of the patients remain functionally dependent.^[36] A reliable prediction model of the patients' outcomes after SAH is needed for decision making of the treatment strategies (clipping with craniotomy,

endovascular treatment, only cerebral ventricular drainage, or conservative therapy) and save the limited medical resources. It is also useful for SAH patients and their families to decide whether or not to receive surgical treatment.

According to the Japanese Guidelines for the Management of Stroke 2015,^[46] the surgical indication was determined mainly depending on the Hunt and Hess grade,^[14] Hunt and Kosnik grade,^[15] or World Federation of Neurosurgical Societies (WFNS) grade.^[34] However, not only the neurological severity grade but also patients' age, locations, sizes of the aneurysms, and their comorbidities have large effects on the outcomes,^[25,26,38,40] so we should determine the treatment strategy after a comprehensive evaluation.

Previously, many studies tried to make the prediction model for SAH outcomes,^[1,13,19,20,37,49,50] and their area under the curve (AUC) of the receiver operating curve varied from 0.700 to 0.898. However, the use of these models is a bit difficult for clinical application because they need some poorly definable variables such as hypertension, SAH thickness, and detailed physiologic scores. In 2019, Donkelaar *et al.* reported SAFIRE score^[3] for prediction of SAH outcomes using only four items, which are easy and relatively objective; age, WFNS grade,^[34] size of the aneurysm, and Fisher computed tomography (CT) scale [Table 1].^[5] They used retrospective data of their single-center cohort of 1215 SAH patients for the training dataset, prospective 224 SAH patients for temporary validation, and performed external validation using 2143 patients from the International Subarachnoid Aneurysm Trial database.^[32] Then, the SAFIRE scoring system achieved the AUC of 0.90 for the temporary validation and that of 0.73 for the external validation.

Similar to these studies, statistically making a prediction model or scoring system need a large number of samples over thousands, so these studies tend to be country-initiated or academic association-initiated research. However, the larger the sample size, the less detailed information is available, such as comorbidities, use of antithrombotic drugs, or laboratory test data and the more there are missing data. Furthermore, the treatment strategies vary from hospital to hospital, and patient backgrounds differ depending on countries and regions. Therefore, these prediction models work as the greatest common denominator worldwide, but not necessarily applicable, with high accuracies, to the respective hospital.

Recently, artificial intelligence (AI) is attracting. Especially, it is a transitional period regarding AI from machine learning to deep learning (DL). Machine learning, such as random forest, logistic regression, or clustering, is defined as "Algorithms that parse data, learn from that data, and then apply what they have learned to make informed decisions." On the other hand, DL is considered an evolution of machine learning. DL uses a programmable neural network that

Table 1: SAFIRE score.^[3]

Variables	Points
Size of the aneurysm (mm)	
<10	0
10–19.9	2
≥20	6
Age	
≤50 y.o.	0
50–60 y.o.	1
60–70 y.o.	2
≥70 y.o.	5
Fisher grade	
1–3	0
4	2
WFNS grade	
I	0
II	2
III	3
IV	6
V	9
Total SAFIRE score	Risk of mRS 4–6 at 2 months*
0–2	<10%
3–5	10–25%
6–8	25–50%
9–15	50–90%
15–22	>90%

mRS: Modified Rankin Scale, WFNS: World Federation of Neurosurgical Societies, y.o.: Years old, *in this study, we did not use these probability but the total SAFIRE score itself and evaluate the association of the outcomes and the total SAFIRE score ranging from 0 to 22

enables machines to make accurate decisions without help from humans. To achieve this, DL applications use a layered structure of algorithms called an artificial neural network (ANN). The design of ANN is inspired by the biological neural network of the human brain, leading to a process of learning that's far more capable than that of standard machine learning models.

Machine learning has been used in neurosurgical situations,^[27,33,39,42,47] but gradually DL is starting to be used as well in decision making for spinal canal stenosis,^[2] predicting outcomes after stroke,^[6] detecting seizure in intracranial electroencephalography recordings (UPenn and Mayo Clinic's Seizure Detection Challenge),^[44,48] pathological diagnosis,^[29] or radiomics studies of brain tumors.^[7,30] However, regarding predicting outcomes of SAH, prediction models only using random forests, categorized as machine learning, were made^[39,47] with an accuracy of 70.9% from the 147 patients^[39] or AUC of 0.837 from the 441 patients,^[47] and there were no reports on the prediction model for SAH outcomes using DL. We hypothesized that we could make a good prediction model for our own hospital using DL, even with a small dataset. Therefore, we herein produced the prediction model using DL software, Prediction One (Sony

Network Communications Inc., Tokyo, Japan)^[41] with our SAH dataset and compared the utility of the model made by Prediction One to SAFIRE score for clinical application.

MATERIALS AND METHODS

Study population

We retrospectively retrieved data from medical records of all the consecutive 153 aneurysmal SAH patients who were admitted between 2012 and 2019 and treated at our institution. Patients with cardiopulmonary arrest on arrival were excluded from the study. The diagnosis of SAH was based on the clinical history and the presence of SAH on CT. The hospital's research ethics committee approved this study, and we gained written informed consent for this study from all of the patients, the legally authorized representative of the patients, or next of kin of the deceased patients. All methods were carried out in accordance with relevant guidelines and regulations (Declaration of Helsinki).

General management of SAH was similar in all cases: all patients were first treated with nicardipine and kept normovolemic with normal blood pressure (systolic blood pressure <140 mmHg). Indication for surgery was established according to the Japanese Guidelines for the Management of Stroke 2009^[45] and 2015.^[46] Both versions describe samely as follows; patients classified as Grade I-III in each grading system^[14,15,34] were considered eligible to undergo aneurysm treatment, whereas those with Grades IV and V were not regarded as suitable for such treatment, except for young and middle-aged patients or patients with large intra-parenchymal hematoma or hydrocephalus. We use WFNS grade on admission, but not Hunt and Hess nor Hunt and Kosnik grade. We mainly performed clipping, but endovascular coiling was considered when it seemed superior to clipping, such as in the posterior circulation aneurysm. Patients with severe hydrocephalus and WFNS Grade IV or V underwent cerebral ventricular drainage, and clipping or coiling was performed when their neurological status improved. Other patients who were not suitable for surgical treatment were treated conservatively. Patients with SAH due to trauma, arteriovenous malformation, or dissection were excluded from this study. Clipping or coiling was performed within 72 h after onset.

All SAH patients who underwent aneurysm clipping or coiling received fasudil, cilostazol, and statin as appropriate after the operation. Rehabilitation of 150 days as maximum and nutritional support was started as soon as possible after the operation, and prophylaxis and treatment of complications were also ensured. Intra-arterial infusion of fasudil was performed when necessary for the treatment of symptomatic vasospasm. In addition, a ventriculoperitoneal shunt was performed when hydrocephalus was observed.

Clinical variables

We collected data regarding physiological symptoms at admission for patients included in this study, that is, age, sex, height, weight, WFNS grade, systolic blood pressure, administration of antithrombotic drugs, history of smoking and drinking, hypertension, diabetes mellitus, and dyslipidemia. We also measured albumin, white blood cell, lymphocyte, triglycerides, total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, glucose, hemoglobin A1c, and brain natriuretic peptide levels at admission. Albumin, lymphocyte, and total cholesterol are known factors for controlling nutritional status score to assess the nutritional status of the patients.^[17]

We determined the location (anterior cerebral artery, anterior communicating artery, internal carotid artery, middle cerebral artery, posterior cerebral artery, basilar artery, vertebral artery, or undetermined due to severe neurological status treated by conservative therapy as well as diffuse SAH with multiple aneurysms which we could not be judged as ruptured) and size (mm) of the aneurysm, Fisher CT scale,^[5] temporal muscle thickness (mm), and area (mm²) as an indicator of systemic skeletal muscle mass^[8-10,22-24,26,43] based on the results of the CT or CT angiography at admission. We used the Aquilion ONE (Canon Medical Systems Corporation, Tochigi, Japan) to take CT and CT angiography images of 0.5 × 0.5 × 1.0 mm voxels. The slice thickness was reconstructed to 5 mm. The window width was adjusted to 90 and the window level to 40 for Fisher CT scale evaluation, while window width was adjusted to 300 and the window level to 20 for temporal muscle measurement. Volume rendered images of the CT angiography were made using Ziostation2 (Ziosoft, Tokyo, Japan). SYNAPSE V 4.1.5 imaging software (Fujifilm Medical, Tokyo, Japan) was used for measurement of the aneurysm size and Fisher CT scale. Temporal muscle area and thickness were also measured using SYNAPSE V through the methods described previously.^[24,26]

We also investigated the treatment strategy (clipping, coiling, or others, including cerebral ventricular drainage or conservative therapy). To evaluate the outcomes, mRS scores at 6 months after the treatment of all 153 patients were collected by either personal outpatient interviews, reports from the rehabilitation hospital or home doctor, or interviews over the telephone, once the ethical approval was obtained for the study. We dichotomized mRS scores into favorable (mRS 0–3) or poor (mRS 4–6).

Making prediction model by Prediction One

We used Prediction One software to make the prediction model. We divided our 153 patients data randomly into 102 patients training dataset and 51 patients external validation dataset. Prediction One read the 102 patients

data and automatically divided them into almost half as internal training and cross-validation datasets. Prediction One automatically adjusted and optimized the variables in a way that is easy to process statistically and mathematically, and select appropriate algorithm with ensemble learning. The missing values were automatically compensated and Prediction One made the best prediction model by ANN with internal cross-validation. The details are trade secrets and could not be provided.

We let the Prediction One software make 2 prediction models using 102 patients training dataset. One was made using all 28 variables acquired at admission, and the other used only four variables which are used for SAFIRE score;^[3] age, WFNS grade, size of the aneurysm, and Fisher CT scale. The AUC of each model and strong variables were automatically calculated. Then, we performed external validation using the 51 patients datasets and calculated AUC, accuracy, precision, recall, and F value, which were used for the evaluation of the prediction model made by AI.

Prediction using SAFIRE score

As the third model in this study, we also investigated SAFIRE scores^[3] and evaluated AUCs using the same 102 patients of the training dataset and 51 patients of external validation dataset, respectively. However, due to some missing data (size of the aneurysm or Fisher CT scale), we calculated the SAFIRE score using 95 of the 102 patients and 46 of the 51 patients datasets, respectively.

SAFIRE score is a simple scoring system predicting the probability of the outcomes as mRS 0–3 or 4–6 at 2 months. In the original article of the SAFIRE score,^[3] SAFIRE score was calculated from the regression equation based on the β coefficients of the regression analyses using age, WFNS grade, size of the aneurysm, and Fisher CT scale. However, it is a complicated equation, so they turned it into a simple scoring system and confirmed its reliability. After the calculation of the sum score, in the original article, the total SAFIRE score was categorized into five groups according to the probability of the outcomes [Table 1]. We similarly calculated the total SAFIRE score from the scoring system, but investigated the association between the outcomes at 6 months and the raw total SAFIRE score itself ranging from 0 to 22. Its AUC was calculated and compared it with those from Prediction One models.

Statistical analysis

Results are shown as median (interquartile range). The difference between the training dataset and the external validation dataset was tested using the Mann–Whitney U test, Fisher's exact test, or Pearson's Chi-square test. A two-tailed $P < 0.05$ was considered statistically significant. We

calculated AUCs and their P -values using SPSS software version 24.0.0 (IBM, New York, USA).

RESULTS

Clinical characteristics

The clinical characteristics of the 153 SAH patients (100 women and 53 men) are summarized in [Table 2]. The median (interquartile range) age was 67 (57–76), WFNS grade 2 (2–4), aneurysm size 6.0 (4.0–8.0) mm, and Fisher group 3 (3–3). Clipping was performed for 121 patients, coiling for 12 patients, and cerebral ventricular drainage or conservative therapy was performed for 20 patients. The median mRS was 2 (0–5) at 6 months and 96 patients (63%) were independent in ADLs. There were no significant differences in the variables between the training and validation data sets.

Model development and validation

Prediction One produced each prediction model in <1 min. The AUCs of each model were described in [Table 3]. The model made by Prediction One using 28 variables had AUC of 0.848 and F value of 0.723, and its AUC for the validation cohort was 0.953 (95%CI 0.900–1.000) with 90% accuracy. The model made by Prediction One using four variables had AUC of 0.803 and F value 0.704, and its AUC for the validation cohort was 0.977 (95%CI 0.938–1.000) with 88% accuracy. The accuracy, precision, and recall were 0.745, 0.630, and 0.850 in the model using all 28 variables, and those in the model using the four variables were 0.745, 0.646, and 0.775, respectively.

The stronger variables of each model are listed in [Table 4]. In the model using the 28 variables, WFNS grade, treatment strategy (conservative therapy was related to poor outcome), size of the aneurysm, temporal muscle area, weight, height, glucose level, systolic blood pressure, triglycerides level, and lymphocyte count had large effects on the outcomes. While in the model using the four variables, WFNS grade, size of the aneurysm, age, and Fisher CT scale were important, in order.

Comparison to SAFIRE score

We calculated the SAFIRE score using 95 of the 102 patients in the training dataset and 46 of the 51 patients in the validation datasets, respectively. The AUCs were 0.875 (95% CI 0.807–0.943) and 0.960 (95% CI 0.905–1.000), respectively [Table 3].

DISCUSSION

We made prediction models using DL software, Prediction One, and we created models with a high prediction rate using

Table 2: Characteristics of the datasets.

Variables	Total (n=153)	Training dataset (n=102)	External validation dataset (n=51)	P-value*
Age (years)	67 (57–76)	67 (57–76)	68 (59–76)	0.360
32–50	21 (14%)	17 (17%)	4 (8%)	
51–65	53 (35%)	33 (32%)	20 (40%)	
66–75	38 (25%)	26 (25%)	12 (23%)	
76–85	28 (18%)	20 (20%)	8 (15%)	
86–96	13 (8%)	6 (6%)	7 (14%)	
Women:Men (%Women)	100:53 (65%)	63:39 (62%)	37:14 (73%)	0.211
Height (cm) (n=143)	155 (150–165)	155 (150–164) (n=96)	155 (150–167) (n=47)	0.938
Weight (kg) (n=142)	52 (45–60)	52 (44–60) (n=95)	52 (47–60) (n=47)	0.857
WFNS grade	2 (2–4)	2 (2–4)	2 (2–4)	0.449
Grade I	29 (19%)	18 (18%)	11 (22%)	
Grade II	58 (38%)	37 (36%)	21 (40%)	
Grade III	8 (5%)	7 (7%)	1 (2%)	
Grade IV	24 (16%)	17 (17%)	7 (14%)	
Grade V	34 (22%)	23 (22%)	11 (22%)	
Aneurysm location				0.469
ACA	17 (11%)	12 (12%)	5 (10%)	
ACoA	36 (24%)	22 (22%)	14 (27%)	
ICA	37 (24%)	24 (24%)	13 (25%)	
MCA	41 (27%)	30 (28%)	11 (22%)	
PCA	4 (3%)	4 (4%)	0	
BA	6 (4%)	2 (2%)	4 (8%)	
VA	5 (3%)	3 (3%)	2 (4%)	
Undetermined†	7 (4%)	5 (5%)	2 (4%)	
Aneurysm size (mm) (n=140)	6.0 (4.0–8.0)	6.0 (4.1–8.1) (n=94)	5.5 (3.8–8.0) (n=46)	0.630
Fisher group	3 (3–3)	3 (3–3)	3 (3–3)	0.905
Group 1	2 (1%)	2 (2%)	0	
Group 2	18 (12%)	13 (13%)	5 (10%)	
Group 3	103 (67%)	65 (63%)	38 (75%)	
Group 4	30 (20%)	22 (22%)	8 (15%)	
Treatment				0.422
Clipping	121 (79%)	83 (81%)	38 (75%)	
Coiling	12 (8%)	6 (8%)	6 (12%)	
Others	20 (13%)	13 (11%)	7 (13%)	
mRS 6-mo post-op	2 (0–5)	2 (0–5)	2 (0–5)	0.625
mRS 0–3	96 (63%)	62 (61%)	34 (67%)	
mRS 4	12 (8%)	9 (9%)	3 (6%)	
mRS 5	16 (10%)	12 (11%)	4 (8%)	
mRS 6	29 (19%)	19 (19%)	10 (19%)	
History				
History of smoking (n=140)	43/140 (31%)	31/94 (33%)	12/46 (26%)	0.442
History of drinking (n=136)	26/136 (19%)	18/91 (20%)	8/45 (18%)	0.999
Hypertension (n=142)	79/142 (56%)	51/94 (54%)	28/48 (58%)	0.722
Diabetes mellitus (n=139)	16/139 (12%)	9/93 (10%)	7/46 (15%)	0.399
Dyslipidemia (n=141)	21/141 (15%)	12/94 (13%)	9/47 (19%)	0.326
Antithrombotic drugs (n=144)	13/144 (9%)	8/96 (8%)	5/48 (10%)	0.760
Systolic blood pressure on admission (mmHg) (n=134)	156 (139–180)	155 (140–180) (n=89)	157 (137–177) (n=45)	0.393
Laboratory data				
Total protein (mg/dL) (n=128)	7.2 (6.7–7.5)	7.1 (6.7–7.4) (n=85)	7.3 (6.8–7.6) (n=43)	0.386
Albumin (mg/dL) (n=143)	4.2 (3.9–4.5)	4.2 (3.9–4.5) (n=96)	4.2 (4.0–4.6) (n=47)	0.654
White blood cell (/μL) (n=141)	9250 (7440–12340)	9320 (7485–12250) (n=95)	8760 (7365–12375) (n=46)	0.464

(Contd...)

Table 2: (Continued).

Variables	Total (n=153)	Training dataset (n=102)	External validation dataset (n=51)	P-value*
Lymphocyte (/μL) (n=141)	1923 (1088–3907)	2270 (1170–4002) (n=95)	1839 (952–2851) (n=46)	0.262
Triglycerides (mg/dL) (n=97)	101 (73–146)	110 (72–170) (n=68)	92 (76–117) (n=29)	0.233
Total cholesterol (mg/dL) (n=86)	192 (171–221)	192 (176–220) (n=60)	185 (167–222) (n=26)	0.628
High-density lipoprotein cholesterol (mg/dL) (n=71)	60 (50–70)	59 (49–71) (n=47)	62 (55–70) (n=24)	0.593
Low-density lipoprotein cholesterol (mg/dL) (n=84)	115 (103–133)	117 (104–132) (n=57)	109 (98–132) (n=27)	0.321
Glucose (mg/dL) (n=111)	158 (134–200)	155 (134–183) (n=72)	170 (134–203) (n=39)	0.513
Hemoglobin A1c (%) (n=90)	5.8 (5.5–6.1)	5.8 (5.5–6.1) (n=59)	5.7 (5.5–6.1) (n=31)	0.865
Brain natriuretic peptide (pg/mL) (n=12)	62.5 (39.9–101.0)	57.9 (32.9–121.1) (n=10)	74.6 (67.0–82.1) (n=2)	0.758
TMT (mm) (n=150)	5.4 (4.2–6.7)	5.3 (4.2–6.8) (n=101)	5.6 (4.5–6.6) (n=49)	0.784
TMA (mm ²) (n=148)	249 (149–366)	242 (146–364) (n=99)	235 (155–365) (n=49)	0.712

ACA: Anterior cerebral artery, ACoA: Anterior communicating artery, BA: Basilar artery, ICA: Internal carotid artery, MCA: Middle cerebral artery, mRS 6-mo post-op: Modified Rankin Scale 6 months after the operation, PCA: Posterior cerebral artery, TMA: Temporal muscle area, TMT: Temporal muscle thickness, VA: Vertebral artery, WFNS grade: World Federation of Neurosurgical Societies, *Mann–Whitney U test, Fisher's exact test, or Pearson's Chi-square test was performed, †Severe neurological status treated by conservative therapy as well as diffuse SAH with multiple aneurysms which we could not be judged as ruptured

a small dataset with several missing data, and it would be reliable for the prediction in our own hospital. Furthermore, this is the first report on creating a prediction model of SAH using DL.

Advantages of DL

DL is now widely used not only in medicine but also in economics, sociology, and logistics. DL is used by a variety of companies with few variables and data, such as forecasting store stocking, predicting insurance policy renewals, and predicting stock prices and land values. Although the sample size in medical situations may be small compared to them, there are many chronological, cross-sectional, and non-numeric variables, such as chief complaint, medical history, family history, blood test data, vital signs, sentences in the electronic medical record themselves, and radiological imaging data with detailed radiomics, that may be suitable for creating predictive models with DL.

Conventional time and cost-consuming statistical analysis need variable optimization like a logarithmic transformation for making the variables like the normal distribution to increase the accuracy of the prediction model. It also requires the arbitrary selection of variables based on previous studies and multivariate analysis needs 10 folds number of samples against the variables.^[35] Therefore, there is a risk that variables which might be important cannot be included in the statistical analysis, or that even the multivariate analysis cannot be performed in a small hospital with small data. Furthermore, in statistical analysis, when there are missing data, we should do multiple imputation or Listwise deletion, which also affects accuracy.

However, DL has the potential to overcome these problems. In the statistical analysis, we should perform variable optimization and sometimes choose variables arbitrarily due to small sample size. However, DL can develop useful prediction models without those time-consuming or arbitrary procedures because DL software automatically does these processes. Furthermore, the number of the variables for DL software is not limited, and DL sometimes find interesting variables as important that has not been taken into account in the previously reported statistical models.^[30,39] Furthermore, DL software automatically substitutes appropriate values instead of the missing ones and calculate the best prediction model without our statistical trial and error.

We then review these benefits of DL in our study. Conventionally, we could have used only ten variables for statistical analysis due to the small sample size of the training dataset ($n = 102$). Furthermore, the dataset contains several missing data. However, we could use 28 variables for making a prediction model by Prediction One, and make a good prediction model from the small dataset. We did not need to perform variable optimization nor manipulations for the missing values. Furthermore, some of the serological test results like glucose level were judged to be important among many other previously reported important factors such as aneurysm location and Fisher CT scale. Besides, the time needed for creating each model was <1 min. Finally, the models achieved high accuracy with AUC of 0.848 in the training dataset and 0.953 in the validation dataset. Putting it bluntly, this means our prediction model, even made from the small dataset, can predict the SAH patients' outcomes treated in our hospital with as high accuracy of SAFIRE score, which was made from the large cohort study.

Table 3: Models for prediction.

Model	AUC derived from the training cohort (n=102)	F value	AUC derived from the validation cohort (n=51)	Accuracy for the validation cohort (%)
Prediction One using 28 variables	0.848 [†]	0.723 [†]	0.953 (95% CI 0.900–1.000)*	90.2
Prediction One using 4 variables [‡]	0.803 [†]	0.704 [†]	0.977 (95% CI 0.938–1.000)*	88.2
SAFIRE score using 4 variables [‡]	0.875 (95%CI 0.807–0.943)* (n=95)	-	0.960 (95%CI 0.905–1.000)* (n=46)	-

AUC: Area under the curve, *P<0.001 for the receiver operating curve calculated by SPSS software, [†]Automatically calculated by Prediction One, [‡]Four variables include age, World Federation of Neurosurgical Societies grade, size of the aneurysm, Fisher computed tomography scale

Table 4: Stronger variables of each model made by Prediction One.

Variables; order of strength	Prediction One using 28 variables	Prediction One using 4 variables
1	WFNS grade	WFNS grade
2	Treatment strategy	Size of the aneurysm
3	Size of the aneurysm	Age
4	Temporal muscle area	Fisher CT scale
5	Weight	
6	Height	
7	Glucose level	
8	Systolic blood pressure	
9	Triglycerides level	
10	Lymphocyte count	

CT: Computed tomography, WFNS: World Federation of Neurosurgical Societies

Limitations of DL

First, the adequate sample size is still unknown for DL. Fujita *et al.* reported that there were no differences in the accuracy of the predicting independent dressing on month after stroke among 80, 100, and 120 patients dataset. However, the accuracy was worse under 60 patients dataset.^[6] Furthermore, an appropriate number of the variables to use to DL is also unestablished. In general, the more variables are used for DL, the higher the accuracy becomes. However, our model using four variables is superior in AUC to that using 28 variables in the validation dataset. This is because some variables may work as noise or have too many missing values to work appropriately. Furthermore, the quality of the data itself may be poor, such as items measured manually, qualitative, or subjective. Furthermore, we should perform dimension reduction for clinical use. For example, our model using 28 variables had temporal muscle area, weight, glucose, and triglyceride levels, as well as lymphocyte count as important variables [Table 4]. However, when so many items are shown to be important, it is hard to know which ones to rely on. They seem like indicators of nutrition or aging, so “age” could represent these variables as a unified item. Therefore, we

need to consider which of the 28 items are really needed and to perform dimension reduction.

Second, DL can treat images or sentences, and this is a very strong point compared to other AI algorithms, but we did not use these advantages. We could have used CT images or descriptions themselves in the electronic medical record to making prediction models. Besides, the number of variables for DL is unlimited so, we could have used more variables that are not reported previously such as other laboratory test data, chronological changes of vital signs, meteorological conditions, and ore calendarial factors. Further study or programming should be needed.

Third, the prediction model derived from our own data cannot be applied to other institutions, and the training and validation dataset must be updated to keep up with advances in medical science and changes in surgical techniques. Creating an AI-based prediction model that can be used universally at any hospital will still require country-initiated or academic association-initiated collaborative research at many institutions and may require the same amount of effort as the traditional statistical model creation.

Fourth, we need to examine the clinical usefulness of the prediction model prospectively. For example, by performing surgery only to patients with a good prognosis predicted by the model, we would evaluate the reduction of the workload and stress of the medical staff, and the medical resources and costs could be saved or not. Furthermore, we should examine the changes of the families’ minds in decision making to treat the patients whose outcomes are empirically difficult to predict, such as elderly patients with SAH Grade III, after showing the result of AI prediction.

Fifth, we used Prediction One software, but there are many AI software (frameworks) worldwide, and there are a thousand different ways to assemble the ANN (libraries). Prediction one is suitable for predicting binomial, ordinal, or continuous variables and can treat Japanese sentences themselves. Furthermore, when there are missing values, it automatically compensates. However, the details of how the neural network is assembled and tuned have not been released, so we need to think carefully about the accuracy of

the models and why the variables were judged as important, considering the clinical meaning.

Future outlook

Despite this easiness, advantages, and future potential of DL, the majority of medical staff cannot treat DL software. Staartjes *et al.* reported this is because of lack of skilled resources (staff, equipment) to develop a model, time limitations restricting AI application in clinical practice, lack of AI models for the indications of interest, uncertainty concerning which processes may benefit most from the application of AI algorithms, lack of data to develop a model, and lack of personal conviction of the added value of this new technology.^[42] As simple DL software is being developed, there is a need for an active interest in using it for the benefit of medical staff and patients. Our study is just one example but suggested the utility of DL software. DL-based tailor-made and efficient medicine, depending on each patient and hospital, would be performed as DL software becomes more popular.

For example, in Japan, a nation-wide study revealed that 77% of the 5344 SAH patients underwent clipping and others coiling from 1999 to 2012,^[28] but clipping is performed for <60% of the SAH patients in the United States.^[31] We perform clipping as first-choice treatment, but coiling is performed in other hospitals.^[24] Similarly, postoperative management against vasospasm delayed cerebral ischemia as well as nutrition therapy and rehabilitation vary depending on each hospital's policy and health-care system worldwide. Although there are these differences in the treatment strategies, DL software can produce predicting models specific to individual centers that would be based on their own unique experience in managing SAH patients. Furthermore, with modern electronic medical records, the clinical variables and clinical outcome data could be automatically fed to the DL software, leading to progressive improvement in predictions over time. This evolutionary prediction will be a benefit to patients, health-care providers, and hospital managers.

Furthermore, the big data have been stored, such as coronavirus disease 2019 Public Datasets,^[12] Miyagi medical and welfare information network,^[16] Tohoku Medical Megabank,^[11,21] Japanese Stroke Databank,^[28] or the Japan Neurosurgical Database.^[18] When these data are open for researchers, it will spur competition in the development of further prognostic models using such big data, like Kaggle competition.^[48]

Limitation of this study

We used WFNS grade at admission, but the SAFIRE score used the WFNS score assessed after neurological resuscitation^[3,4] (rWFNS; e.g., cerebral spinal fluid drainage

for acute hydrocephalus or evacuation of an intracerebral hematoma). Furthermore, the SAFIRE score predicts 2-month outcomes, but our models those 6 months. These are differences, so simply comparing their AUCs requires caution.

CONCLUSION

We easily and quickly made prediction models using Prediction One software. The accuracies of the prediction models were not inferior to those of previous statistically calculated prediction models. Even with a small single-center dataset, containing missing data, prognostic models made by DL software can be useful at the institution and may be applied to daily clinical practice in the future.

Declaration of patient consent

The authors certify that they have obtained all appropriate patient consent.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Abulhasan YB, Alabdulraheem N, Simoneau G, Angle MR, Teitelbaum J. Mortality after spontaneous subarachnoid hemorrhage: Causality and validation of a prediction model. *World Neurosurg* 2018;112:e79-811.
2. Azimi P, Mohammadi HR, Benzel EC, Shahzadi S, Azhari S. Use of artificial neural networks to decision making in patients with lumbar spinal canal stenosis. *J Neurosurg Sci* 2017;61:603-11.
3. Donkelaar CE, Bakker NA, Birks J, Veeger NJ, Metzemaekers JD, Molyneux AJ, *et al.* Prediction of outcome after aneurysmal subarachnoid hemorrhage: Development and validation of the SAFIRE grading scale. *Stroke* 2019;50:837-44.
4. Donkelaar CE, Bakker NA, Veeger NJ, Uyttenboogaart M, Metzemaekers JD, Eshghi O, *et al.* Prediction of outcome after subarachnoid hemorrhage: Timing of clinical assessment. *J Neurosurg* 2017;126:52-9.
5. Fisher CM, Kistler JP, Davis JM. Relation of cerebral vasospasm to subarachnoid hemorrhage visualized by computerized tomographic scanning. *Neurosurgery* 1980;6:1-9.
6. Fujita T, Ohashi T, Yamane K, Yamamoto Y, Sone T, Ohira Y, *et al.* Relationship between the number of samples and the accuracy of the prediction model for dressing independence using artificial neural networks in stroke patients. *Jpn J Compr Rehabil Sci* 2020;11:28-34.
7. Fukuma R, Yanagisawa T, Kinoshita M, Shinozaki T, Arita H, Kawaguchi A, *et al.* Prediction of IDH and TERT promoter mutations in low-grade glioma from magnetic resonance

- images using a convolutional neural network. *Sci Rep* 2019;9:20311.
8. Furtner J, Berghoff AS, Albtoush OM, Woitek R, Asenbaum U, Prayer D, *et al.* Survival prediction using temporal muscle thickness measurements on cranial magnetic resonance images in patients with newly diagnosed brain metastases. *Eur Radiol* 2017;27:3167-73.
 9. Furtner J, Berghoff AS, Schöpf V, Reumann R, Pascher B, Woitek R, *et al.* Temporal muscle thickness is an independent prognostic marker in melanoma patients with newly diagnosed brain metastases. *J Neurooncol* 2018;140:173-8.
 10. Furtner J, Genbrugge E, Gorlia T, Bendszus M, Nowosielski M, Gofinopoulos V, *et al.* Temporal muscle thickness is an independent prognostic marker in patients with progressive glioblastoma: Translational imaging analysis of the EORTC 26101 trial. *Neuro Oncol* 2019;21:1587-94.
 11. Fuse N, Sakurai-Yageta M, Katsuoka F, Danjoh I, Shimizu R, Tamiya G, *et al.* Establishment of integrated biobank for precision medicine and personalized healthcare: The Tohoku medical megabank project. *JMA J* 2019;2:113-22.
 12. Google Cloud Platform, COVID-19 Public Datasets; 2020. Available from: https://www.console.cloud.google.com/marketplace/product/bigquery-public-datasets/covid19-public-data-program?_ga=2.6492658.-1999449961.1599963188&pli=1. [Last accessed on 2020 Jun 29].
 13. Hostettler IC, Sebök M, Ambler G, Muroi C, Prömmel P, Neidert MC, *et al.* Validation and optimization of barrow neurological institute score in prediction of adverse events and functional outcome after subarachnoid hemorrhage-creation of the HATCH (hemorrhage, age, treatment, clinical state, Hydrocephalus) score. *Neurosurgery* 2020; Doi: 10.1093/neuros/nyaa316
 14. Hunt WE, Hess RM. Surgical risk as related to time of intervention in the repair of intracranial aneurysms. *J Neurosurg* 1968;28:14-20.
 15. Hunt WE, Kosnik EJ. Timing and perioperative care in intracranial aneurysm surgery. *Clin Neurosurg* 1974;21:79-89.
 16. Ido K, Nakamura N, Nakayama M. Miyagi medical and welfare information network: A backup system for patient clinical information after the great east Japan earthquake and tsunami. *Tohoku J Exp Med* 2019;248:19-25.
 17. Ignacio de Ulíbarri J, González-Madroño A, de Villar NG, González P, González B, Mancha A, *et al.* CONUT: A tool for controlling nutritional status. First validation in a hospital population. *Nutr Hosp* 2005;20:38-45.
 18. Iihara K, Tominaga T, Saito N, Suzuki M, Date I, Fujii Y, *et al.* The japan neurosurgical database: Overview and results of the first-year survey. *Neurol Med Chir (Tokyo)* 2020;60:165-90.
 19. Jaja BN, Cusimano MD, Etmnan N, Hanggi D, Hasan D, Ilodigwe D, *et al.* Clinical prediction models for aneurysmal subarachnoid hemorrhage: A systematic review. *Neurocrit Care* 2013;18:143-53.
 20. Jaja BN, Saposnik G, Lingsma HF, Macdonald E, Thorpe KE, Mamdani M, *et al.* Development and validation of outcome prediction models for aneurysmal subarachnoid haemorrhage: The SAHIT multinational cohort study. *BMJ* 2018;360:j5745.
 21. Kamada F, Aoki Y, Narisawa A, Abe Y, Komatsuzaki S, Kikuchi A, *et al.* A genome-wide association study identifies RNF213 as the first Moyamoya disease gene. *J Hum Genet* 2011;56:34-40.
 22. Katsuki M, Kakizawa Y, Nishikawa A, Yamamoto Y, Uchiyama T. Lower total protein and absence of neuronavigation are novel poor prognostic factors of endoscopic hematoma removal for intracerebral hemorrhage. *J Stroke Cerebrovasc Dis* 2020;29:105050.
 23. Katsuki M, Kakizawa Y, Nishikawa A, Yasunaga Y, Uchiyama T. Endoscopic hematoma removal of supratentorial intracerebral hemorrhage under local anesthesia reduces operative time compared to craniotomy. *Sci Rep* 2020;10:10389.
 24. Katsuki M, Suzuki Y, Kunitoki K, Sato Y, Sasaki K, Mashiyama S, *et al.* Temporal muscle as an indicator of sarcopenia is independently associated with hunt and kosnik grade on admission and the modified rankin scale score at 6 months of patients with subarachnoid hemorrhage treated by endovascular coiling. *World Neurosurg* 2020;137:e526-34.
 25. Katsuki M, Suzuki Y, Sato Y, Sasaki K, Shingai Y, Mashiyama S, *et al.* In reply to the letter to the editor regarding temporal muscle as an indicator of sarcopenia is independently associated with hunt and kosnik grade on admission and the modified rankin scale at 6 months of patients with subarachnoid hemorrhage treated by endovascular coiling. *World Neurosurg* 2020;140:433.
 26. Katsuki M, Yamamoto Y, Uchiyama T, Wada N, Kakizawa Y. Clinical characteristics of aneurysmal subarachnoid hemorrhage in the elderly over 75; would temporal muscle be a potential prognostic factor as an indicator of sarcopenia? *Clin Neurol Neurosurg* 2019;186:105535.
 27. Kim HC, Rhim JK, Ahn JH, Park JJ, Moon JU, Hong EP, *et al.* Machine learning application for rupture risk assessment in small-sized intracranial aneurysm. *J Clin Med* 2019;8:683.
 28. Kobayashi Y. In: Kobayashi Y, editor. *Japanese Stroke Databank* 2015. Tokyo: Nakayama Shoten; 2015.
 29. Kumar R, Gupta A, Arora HS, Pandian GN, Raman B. CGHF: A computational decision support system for glioma classification using hybrid radiomics-and stationary wavelet-based features. *IEEE Access* 2020;8:79440-58.
 30. Lao J, Chen Y, Li ZC, Li Q, Zhang J, Liu J, *et al.* A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci Rep* 2017;7:10353.
 31. Lin N, Cahill KS, Frerichs KU, Friedlander RM, Claus EB. Treatment of ruptured and unruptured cerebral aneurysms in the USA: A paradigm shift. *J Neurointerv Surg* 2012;4:182-9.
 32. Molyneux A, Kerr R, Stratton I, Sandercock P, Clarke M, Shrimpton J, *et al.* International subarachnoid aneurysm trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: A randomised trial. *Lancet* 2002;360:1267-74.
 33. Niftrik CH, der Wouden F, Staartjes VE, Fierstra J, Stienen MN, Akeret K, *et al.* Machine learning algorithm identifies patients at high risk for early complications after intracranial tumor surgery: Registry-based cohort study. *Neurosurgery* 2019;85:E756-64.
 34. Report of world federation of neurological surgeons committee on a universal subarachnoid hemorrhage grading scale. *J Neurosurg* 1988;68:985-6.
 35. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR.

- A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373-9.
36. Rinkel GJ, Algra A. Long-term outcomes of patients with aneurysmal subarachnoid haemorrhage. *Lancet Neurol* 2011;10:349-6.
 37. Risselada R, Lingsma HF, Bauer-Mehren A, Friedrich CM, Molyneux AJ, Kerr RS, *et al.* Prediction of 60 day case-fatality after aneurysmal subarachnoid haemorrhage: Results from the international subarachnoid aneurysm trial (ISAT). *Eur J Epidemiol* 2010;25:261-6.
 38. Rosengart AJ, Schultheiss KE, Tolentino J, Macdonald RL. Prognostic factors for outcome in patients with aneurysmal subarachnoid hemorrhage. *Stroke* 2007;38:2315-21.
 39. Rubbert C, Patil KR, Beseoglu K, Mathys C, May R, Kaschner MG, *et al.* Prediction of outcome after aneurysmal subarachnoid haemorrhage using data from patient admission. *Eur Radiol* 2018;28:4949-58.
 40. Shiue I, Arima H, Hankey GJ, Anderson CS. Location and size of ruptured intracranial aneurysm and serious clinical outcomes early after subarachnoid hemorrhage: A population-based study in Australasia. *Cerebrovasc Dis* 2011;31:573-9.
 41. Sony Network Communications, Prediction One; 2020. Available from: <https://www.predictionone.sony.biz>. [Last accessed on 2020 Mar 29]
 42. Staartjes VE, Stumpo V, Kernbach JM, Klukowska AM, Gadraj PS, Schröder ML, *et al.* Machine learning in neurosurgery: A global survey. *Acta Neurochir (Wien)* 2020. Doi: 10.1007/s00701-020-04532-1.
 43. Steindl A, Leitner J, Schwarz M, Nenning KH, Asenbaum U, Mayer S, *et al.* Sarcopenia in neurological patients: Standard values for temporal muscle thickness and muscle strength evaluation. *J Clin Med* 2020;9:1272.
 44. Temko A, Sarkar A, Lightbody G. Detection of seizures in intracranial EEG: UPenn and mayo clinic's seizure detection challenge. *Annu Int Conf IEEE Eng Med Biol Soc* 2015;2015:6582-5.
 45. The Japan Stroke Society. Japanese Guidelines for the Management of Stroke 2009. Tokyo: Kyowa Kikaku; 2009.
 46. The Japan Stroke Society. Japanese Guidelines for the Management of Stroke 2015. Tokyo: Kyowa Kikaku; 2015.
 47. Toledo P, Rios PM, Ledezma A, Sanchis A, Alen JF, Lagares A. Predicting the outcome of patients with subarachnoid hemorrhage using machine learning techniques. *IEEE Trans Inf Technol Biomed* 2009;13:794-801.
 48. UPenn and Mayo Clinic's Seizure Detection Challenge Detect Seizures in Intracranial EEG Recordings. San Francisco: Kaggle; 2014. Available from: <https://www.kaggle.com/c/seizure-detection>. [Last accessed on 2020 Jun 28]
 49. Witsch J, Frey HP, Patel S, Park S, Lahiri S, Schmidt JM, *et al.* Prognostication of long-term outcomes after subarachnoid hemorrhage: The FRESH score. *Ann Neurol* 2016;80:46-58.
 50. Zeiler FA, Lo BW, Akoth E, Silvaggio J, Kaufmann AM, Teitelbaum J, *et al.* Predicting outcome in subarachnoid hemorrhage (SAH) utilizing the Full Outline of UnResponsiveness (FOUR) score. *Neurocrit Care* 2017;27:381-91.

How to cite this article: Katsuki M, Kakizawa Y, Nishikawa A, Yamamoto Y, Uchiyama T. Easily created prediction model using deep learning software (Prediction One, Sony Network Communications Inc.) for subarachnoid hemorrhage outcomes from small dataset at admission. *Surg Neurol Int* 2020;11:374.