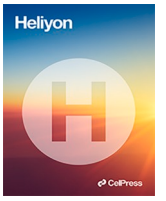
Heliyon 10 (2024) e31000

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Deep learning model integrating radiologic and clinical data to predict mortality after ischemic stroke

Changi Kim [a], Joon-myoung Kwon [b,c,d], Jiyeong Lee [e], Hongju Jo [f], Dowan Gwon [g], Jae Hoon Jang [h], Min Kyu Sung [h], Sang Won Park [i,j], Chulho Kim [k,1,**], Mi-Young Oh [e,*,1]

[a] *Department of Bioengineering, Seoul National University, Seoul, Republic of Korea*
[b] *Medical Research Team, Medical AI Inc, DC, USA*
[c] *Department of Critical Care Emergency Medicine, Incheon Sejong Hospital, Incheon, Republic of Korea*
[d] *Artificial Intelligence and Big Data Research Center, Sejong Medical Research Institute, Bucheon, Republic of Korea*
[e] *Department of Neurology, Bucheon Sejong Hospital, Bucheon, Republic of Korea*
[f] *Independent Researcher*
[g] *Department of Digital&Biohealth, Group of AI/DX Business, KT, Seoul, Republic of Korea*
[h] *Department of Family Medicine, College of Medicine, KyungHee University, Seoul, Republic of Korea*
[i] *Department of Medical Informatics, School of Medicine, Kangwon National University, Chuncheon, Republic of Korea*
[j] *Institute of Medical Science, School of Medicine, Kangwon National University, Chuncheon, Republic of Korea*
[k] *Department of Neurology, Hallym University College of Medicine, Chuncheon, Republic of Korea*

A R T I C L E I N F O

Keywords:
Deep learning
Mortality
Stroke

A B S T R A C T S

Objective: Most prognostic indexes for ischemic stroke mortality lack radiologic information. We aimed to create and validate a deep learning-based mortality prediction model using brain diffusion weighted imaging (DWI), apparent diffusion coefficient (ADC), and clinical factors.
Methods: Data from patients with ischemic stroke who admitted to tertiary hospital during acute periods from 2013 to 2019 were collected and split into training (n = 1109), validation (n = 437), and internal test (n = 654). Data from patients from secondary cardiovascular center was used for external test set (n = 507). The algorithm for predicting mortality, based on DWI and ADC (DLP_DWI), was initially trained. Subsequently, important clinical factors were integrated into this model to create the integrated model (DLP_INTG). The performance of DLP_DWI and DLP_INTG was evaluated by using time-dependent area under the receiver operating characteristic curves (TD AUCs) and Harrell concordance index (C-index) at one-year mortality.
Results: The TD AUC of DLP_DWI was 0.643 in internal test set, and 0.785 in the external dataset. DLP_INTG had a higher performance at predicting one-year mortality than premise score in internal dataset (TD- AUC: 0.859 vs. 0.746; p = 0.046), and in external dataset (TD- AUC: 0.876 vs. 0.808; p = 0.007). DLP_DWI and DLP_INTG exhibited strong discrimination for the high-risk group for one-year mortality.

* Corresponding author. Department of Neurology, Bucheon Sejong Hospital, 28, Hohyeon-ro 489, beon-gil, Bucheon-si, Gyeonggi-do, 14754, Republic of Korea.
** Corresponding author. Department of Neurology, Hallym University College of Medicine, 1 Hallymdaehak-gil, Chuncheon-si, Gangwon-do, 24252, Republic of Korea.
E-mail addresses: gumdol52@naver.com (C. Kim), mohm99@gmail.com (M.-Y. Oh).
[1] These authors contributed equally as a corresponding author.

https://doi.org/10.1016/j.heliyon.2024.e31000
Received 27 November 2023; Received in revised form 8 May 2024; Accepted 9 May 2024
Available online 16 May 2024

Interpretation: A deep learning model using brain DWI, ADC and the clinical factors was capable of predicting mortality in patients with ischemic stroke.

1. Introduction

Stroke is the second-leading cause of mortality worldwide with an annual mortality rate of about 5.5 million. Stroke also superimposed a substantial burden of chronic morbidity necessitating protracted medical attention and institutional care [1]. It is important to predict the life expectancy of patients with stroke in acute period for facilitating treatment stratification and the provision of essential information to caregivers. The previous studies have explored risk stratification systems to predict stroke patient outcomes using the various clinical determinants, including age, stroke severity, and comorbidities such as diabetes, cardiovascular disease, and chronic kidney disease [2]. However, most of these investigations have primarily focused on short-term outcome, with a marked emphasis on clinical variables [1–6]. Moreover, There has been growing evidence of the potential value of radiologic information from brain diffusion-weighted imaging (DWI) and apparent diffusion coefficient (ADC), such as infarction volume, specific lesion location or patterns has been suggested as valuable for mortality prediction in ischemic stroke [7,8]. However, the quantification and detection of these radiological findings from brain DWI and ADC are labor-intensive tasks, and their integration into mortality prediction models has been relatively limited. Mittal et al. reported a scoring system to predict in-hospital mortality after acute stroke primarily revolved around clinical variables, emphasizing the initial neurological severity, measured by National Institute of Health stroke scale (NIHSS) score and also the Glasgow coma scale [3]. The premise score, another scoring system for predicting early mortality after stroke. It composed of important clinical factors such as age, NIHSS score, and itemized imaging variables such as existence of posterior circulation, non-lacunar stroke. Despite being well-developed and validated, it had limitations in using the comprehensive radiological information [4].

Recently deep learning models, specifically convolutional neural networks (CNNs) could be used to extract relevant information from brain DWI and ADC of stroke patients with greater efficiency and accuracy [9,10]. Furthermore, the successful integration of radiologic information from CNNs using brain DWI, ADC, and pertinent clinical factors would have the potential to enhance prognostic capabilities significantly, compared to previous scoring system [11,12]. In line with this hypothesis, we developed and validated a deep learning–driven mortality prediction model for stroke patients. Additionally, we incorporated a time-dependent loss function to account for dynamic contribution of radiologic information during the follow-up period [13]. We initially developed the prediction model using exclusively brain DWI and ADC data, denoted as DLP_DWI, and subsequently integrated this radiological information with important clinical factors to create an integrated predictive model, referred to as DLP_INTG.

2. Methods

2.1. Dataset

This retrospective observational cohort study was designed to develop the prediction model for one-year mortality in stroke patients during the acute periods. This study utilized two different datasets, creating a developmental cohort including training, validation, an internal test set, and a geographically different external validation set. We used an open public dataset as a developmental

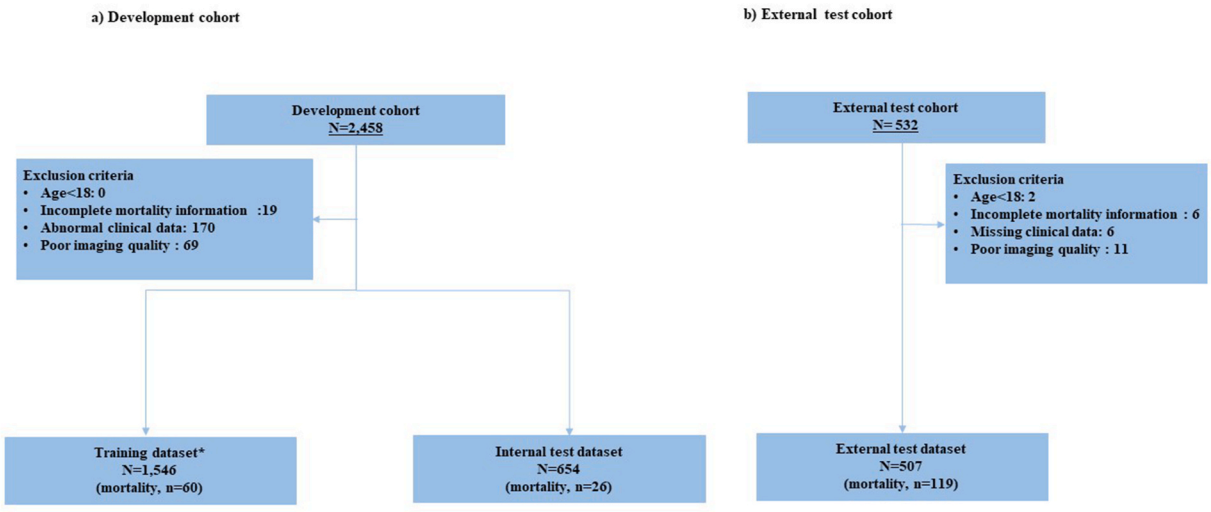


Fig. 1. The flow chart of study population.

cohort, which was collected under the leadership of the Korean government and was available at https://www.aihub.or.kr. The developmental cohort included the stroke patients who admitted to the tertiary hospital between January 2013 and June 2019. The patients in the developmental cohort were split into training, validation, and internal test datasets. Data from patients who admitted to secondary cardiovascular hospital between October 2016 and May 2022 were used as an external validation dataset. The exclusion criteria were as follows: age <18, incomplete mortality information, where the deceased date was not matched, missing or abnormal clinical data that were deemed to be erroneous due to input errors, and poorly qualified imaging data, as shown at Fig. 1.

2.2. Data collection for the prediction model

The variables from the medical records were collected including the following:

1) Demographic information: age, sex, height, weight, body mass index, preexisting morbidities such as hypertension, diabetes, hyperlipidemia, transient ischemic attack or stroke, peripheral artery disease, ischemic heart disease, atrial fibrillation, cancer and premorbid state
2) Initial stroke severity, measured by National Institute of Health stroke scale (NIHSS) score
3) Initial brain DWI and ADC magnetic resonance imaging upon stroke admission before treatment; the detailed information of image acquisition was shown in Supplementary Table S6.
4) Mortality (all-cause mortality) data for the development cohort were obtained from public open data, as were other clinical variables. Mortality data of external test cohort (Bucheon sejong hospital) were obtained from the database of the Ministry of the Interior and Safety, Republic of Korea, and survival time was censored on July 15, 2022.
5) Hyperacute stroke treatment; intravenous or intra-arterial thrombolytic treatment, and combined treatment
6) Initial vital sign and laboratory findings at admission: systolic and diastolic blood pressure, hemoglobin, hematocrit, white blood cell count, platelet count, blood urea nitrogen, creatinine, total cholesterol, high density lipoprotein (HDL) cholesterol, low density lipoprotein (LDL) cholesterol, glucose, prothrombin time

2.3. Development of deep learning mortality prediction model using brain DWI (DLP_DWI)

2.3.1. Neural network

A 3D CNN model for predicting one-year mortality was developed using paired brain DWI and ADC images from patients in the developmental cohort (Fig. 2). To ensure compatibility with the CNN model, all images were resized to 30 × 256 × 256 pixels and pixel values were rescaled to fit the range of zero to 255. Prior to inputting the images into the CNNs, the resized DWI and ADC image pairs were concatenated. For the loss function, we employed a negative log-likelihood and incorporated non-proportional hazards using Nnet-survival [13]. After experimenting with two CNN models, ResNet50, and DenseNet169, the higher-performing model was selected based on the area under the receiver operating characteristic curve (AUC) [10,14]. Subsequently, the last layer of the model

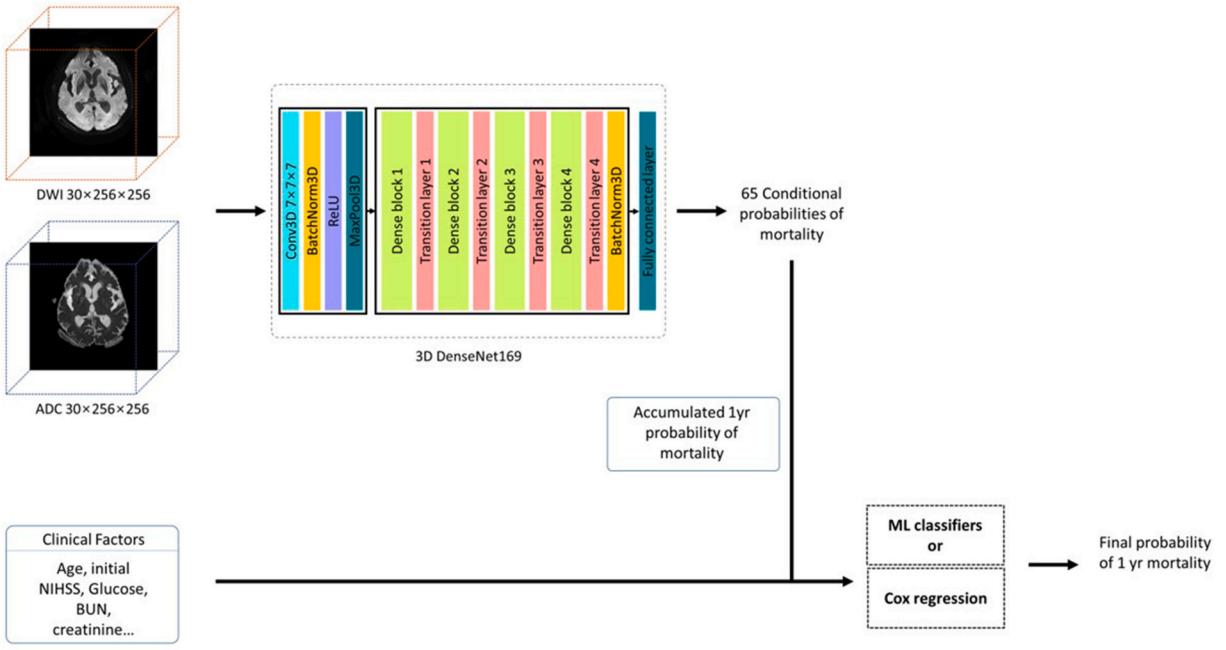


Fig. 2. Architecture of the deep learning based survival prediction model in stroke patients.

was replaced by a fully connected layer of 65 perceptrons, each corresponding to a 7.5-day interval of time. The resulting output consisted of conditional probabilities of survival at different time intervals. The final output was defined as the probability of one-year mortality which served as the primary aim of our study; 1-[cumulative survival probability (product of the consecutive 48 perceptrons)]. To visualize the performance of DLP_DWI, gradient-class activation maps (Grad-CAM) were generated for the third-to-last convolutional layer [15].

2.4. Integration of deep learning prediction model and clinical variables (DLP_INTG)

The architecture of the deep learning-based survival prediction model, which utilizes clinical and radiological information, was shown in Fig. 2. The Selection of clinical factors for inclusion in the integrated model was based on their statistical significance in the Cox regression model as well as their importance as determined by clinicians. We developed an ensemble model to merge with the output of DLP_DWI and selected clinical factors using the ML classifier, such as Random Forest, Catboost, and XGboost [16–18]. Additionally, we derived a Cox regression model based on multivariable proportional hazards using the output of DLP_DWI and clinical factors. Following a comparison of performance between the machine learning classifiers and the Cox regression model, we selected the best-performing model as an integrated model (DLP_INTG), which included the output from DLP_DWI and the essential clinical factors. Subsequently, the performance of DLP_INTG was evaluated and compared with a conventional risk scoring system, premise score.

2.5. Statistical analysis

The discriminative performance of DLP_DWI, DLP_INTG, and the premise score was assessed using time-dependent areas under the receiver operating characteristic curve (TD AUCs) and the Harrell concordance index (C-index) one year after stroke [19,20]. We chose TD AUC as the primary measure to assess prediction model's discriminative performance due to the time-dependent loss function of

Table 1
Baseline characteristics of developmental and external test cohorts according to mortality.

	Developmental cohort			External test cohort		
Characteristic	Mortality (−)	Mortality (+)	*p*	Mortality (−)	Mortality (+)	*p*
	N = 2114	N = 86		N = 388	N = 119	
Age	67.4 ± 12.6	79.0 ± 9.6	0.000	67.9 ± 12.4	79.9 ± 9.9	0.000
Sex			0.484			0.001
Female	890 (42.1 %)	40 (46.5 %)		166 (42.5 %)	68 (57.1 %)	
Male	1224 (57.9 %)	46 (53.5 %)		225 (57.5 %)	51 (42.9 %)	
Body Mass Index, kg/m^2	23.8 ± 3.2	22.3 ± 2.9	0.000	24.5 ± 3.3	22.7 ± 2.5	0.000
NIH Stroke Scale	3.6 ± 4.1	7.2 ± 6.1	0.000	4.7 ± 6.8	14.2 ± 13.4	0.000
Previous Stroke	319 (15.1 %)	22 (25.6 %)	1.000	52 (13.4 %)	30 (25.2 %)	0.008
Hypertension	1258 (59.5 %)	51 (59.3 %)	1.000	252 (64.9 %)	72 (60.5 %)	0.458
Diabetes	590 (27.9 %)	33 (38.4 %)	0.047	134 (34.3 %)	40 (33.6 %)	0.784
Ischemic Heart Disease	122 (5.8 %)	9 (10.5 %)	0.116	84 (21.6 %)	32 (26.9 %)	0.793
Atrial Fibrillation	373 (17.6 %)	33 (38.4 %)	0.000	78 (20.1 %)	30 (25.2 %)	0.288
Smoking History	709 (33.5 %)	21 (24.4 %)	0.100	80 (20.6 %)	25 (21.0 %)	1.000
Hyperlipidemia	482 (22.8 %)	19 (22.1 %)	0.982	128 (32.7 %)	27 (23.3 %)	0.377
Peripheral Arterial Disease	11 (0.5 %)	1 (1.2 %)	0.963	19 (4.9 %)	3 (2.5 %)	0.392
Cancer	9 (0.4 %)	1 (1.2 %)	0.858	12 (3.1 %)	8 (6.7 %)	0.131
Thrombolysis Treatment			0.738			1.000
No	1823 (86.2 %)	73 (84.9 %)		342 (88.1 %)	110 (92.4 %)	
Intravenous Thrombolysis	172 (8.1 %)	7 (8.1 %)		46 (11.9 %)	9 (7.6 %)	
Intraarterial Thrombolysis	57 (2.7 %)	4 (4.7 %)		0 (0.0 %)	0 (0.0 %)	
Combined	62 (2.9 %)	2 (0.3 %)		0 (0.0 %)	0 (0.0 %)	
Hemoglobin, g/dL	13.9 ± 1.9	12.8 ± 2.2	0.000	13.7 ± 1.9	12.6 ± 2.3	0.000
Hematocrit, %	40.6 ± 5.1	37.4 ± 5.8	0.000	40.4 ± 5.1	37.3 ± 6.4	0.000
White Blood Cell count, 10^3/μL	8.1 ± 2.8	8.6 ± 3.6	0.169	8.0 ± 2.6	9.4 ± 5.1	0.004
Platelet, 10^3/μl	224.7 ± 64.0	220.2 ± 84.4	0.627	231.4 ± 66.2	218.1 ± 67.6	0.058
Blood Urea Nitrogen, mg/dL	16.3 ± 6.9	19.8 ± 8.4	0.000	17.3 ± 6.8	24.5 ± 17.4	0.000
Creatinine, mg/dL	0.9 ± 0.8	1.0 ± 0.9	0.183	1.0 ± 0.5	1.3 ± 1.0	0.003
Total Cholesterol, mg/dL	178.8 ± 44.7	161.4 ± 39.1	0.000	167.2 ± 45.0	152.7 ± 43.0	0.002
Triglyceride, mg/dL	118.6 ± 66.2	93.8 ± 40.6	0.000	131.1 ± 76.3	103.9 ± 41.6	0.000
HDL Cholesterol, mg/dL	44.8 ± 12.4	43.8 ± 13.4	0.476	45.6 ± 10.8	45.8 ± 10.7	0.826
LDL Cholesterol, mg/dL	114.6 ± 38.7	99.9 ± 33.0	0.001	96.8 ± 31.6	90.4 ± 35.9	0.063
Glucose, mg/dL	136.8 ± 53.8	146.7 ± 61.5	0.098	149.5 ± 71.8	156.5 ± 87.1	0.429
Systolic Blood Pressure, mmHg	139.7 ± 23.9	135.1 ± 25.7	0.079	144.5 ± 26.8	140.2 ± 28.5	0.135
Diastolic Blood Pressure, mmHg	84.2 ± 14.2	82.5 ± 16.4	0.337	81.1 ± 14.9	80.6 ± 16.6	0.917
Premise score	3.5 ± 1.9	5.8 ± 2.4	0.000	4.0 ± 2.0	6.6 ± 2.4	0.000
Infarction volume, mL	4.5 ± 3.2	6.6 ± 4.0	0.000	16.4 ± 58.5	79.7 ± 18.4	0.000
Posterior circulation stroke	316 (14.9 %)	15 (17.4 %)	0.631	145 (37.4 %)	47 (39.5 %)	0.757

Data are presented as number (%) or mean ± SD.

Nnet [19]. We performed pairwise comparisons of TD AUCs. Survival curves were visualized using the Kaplan-Meier method for both DLP_DWI and DLP_INTG, categorizing patients into high- and low-risk groups based on the optimal threshold determined by the Youden index J [21,22]. Additionally, we evaluated the performance of DLP_DWI and DLP_INTG using sensitivity, specificity, negative predictive value, and positive predictive value. Cox proportional hazard analysis was used as a method to develop an integrated model incorporating the output of DLP_DWI and clinical variables. The training dataset used for Cox regression analysis included both the training dataset and validation datasets for DLP_DWI. Variance inflation factor was evaluated from multiple regression analysis to check multicollinearity. We assumed independent censoring, as mortality information until the fixed censor date for all patients were collected. All analyses were performed using R software (version 3.3.3, R Foundation for Statistical Computing) (libraries "survival," "survminer," "timeROC," "moonBook," "survcomp", "PredictABEL," "Hmisc"). For all statistical analyses, $p < 0.1$ for univariable analysis, and $p < 0.05$ for multivariable analysis was considered to indicate statistical significance.

2.6. Ethical approvals, registrations, and patient consents

This study was approved by the Institutional Review Board (IRB) of Bucheon sejong hospital (BSH 2021-09-005) on September 29, 2021. IRB determined that participant consent was waived in this retrospective study. We tried to follow the tripod guideline [23].

3. Results

3.1. Study population

A total of 2,458 patients who were admitted to a tertiary medical center were included. We excluded the following patients according to the study criteria: incomplete mortality information, where the deceased date was not matched (n = 19), abnormal clinical data due to data entry errors (n = 170), and age <18 years (n = 0). The remaining 2,203 patients were included in the final analysis. The study population was divided into the training dataset (n = 1546) and internal test dataset (n = 654). The external test cohort comprised 507 patients who admitted to a secondary cardiovascular medical center applying the exclusion criteria (Fig. 1).

3.2. Baseline characteristics

In the development cohort, 86 of the 2,203 patients (3.9 %) had deceased, while in the external test cohort, 119 of 507 patients (23.5 %) had deceased. In the developmental cohort, deceased patients were more likely to be older, have atrial fibrillation, possess a lower body mass index, and exhibit higher levels of blood urea nitrogen and creatinine compared to surviving patients. Deceased patients in the external test cohort were also more likely to be older, have a higher NIHSS score, compared to surviving patients (Table 1). The proportion of deceased patients in the external test cohort was higher than that in the developmental cohort. Patients in the external test cohort were older and exhibited a higher prevalence of comorbidities such as hypertension, diabetes, hyperlipidemia,

Table 2
Performance of DLP_DWI and DLP_INTG.

	Developmental cohort				External test cohort	
	Training set		Internal test set		External test set	
	TD-AUC	C-index	TD-AUC	C-index	TD-AUC	C-index
As a single predictor						
DLP_DWI*	0.946 (0.901,0.992)	0.875 (0.845,0.905)	0.643 (0.524,0.761)	0.765 (0.704,0.826)	0.785 (0.729,0.842)	0.741 (0.717,0.765)
Infarction volume	0.658 (0.579,0.737)	0.672 (0.593,0.751)	0.628 (0.515,0.741)	0.641 (0.526,0.756)	0.668 (0.594,0.742)	0.657 (0.606,0.710)
Posterior circulation stroke	0.543 (0.478,0.584)	0.597 (0.450,0.745)	0.532 (0.478,0.585)	0.650 (0.321,0.977)	0.517 (0.459,0.582)	0.524 (0.425,0.623)
As an integrated model						
DLP_INTG**	0.955† (0.925,0.984)	0.943 (0.911,0.975)	0.859† (0.777,0.941)	0.852 (0.772,0.932)	0.876† (0.828,0.925)	0.842 (0.630,0,746)
Premise score	0.747 (0.680,0.815)	0.775 (0.706,0.843)	0.746 (0.683,0.813)	0.791 (0.747,0.835)	0.808 (0.756,0.719)	0.780 (0.760,0.800)

*DLP_DWI was superior to the conventional imaging marker, infarction volume. The time-dependent areas under the receiver operating characteristic curve (TD AUCs) of DLP_DWI showed superior performance, compared to the TD AUCs of infarction volume in the training set ($p < 0.001$), in external test set($p = 0.018$). TD AUCs of DLP_DWI was superior to the performance of the specific ischemic location, posterior circulation stroke in the training set ($p < 0.001$), in external test set ($p < 0.001$).

**DLP_INTG is the integrated Cox prediction model which was composed of DLP_DWI, age, NIH stroke scale, glucose, smoking history, atrial fibrillation, previous stroke, hypertension, diabetes, blood urea nitrogen (BUN), and creatinine. The training dataset used for Cox regression analysis included both the training dataset and validation datasets for DLP_DWI. †The time-dependent areas under the receiver operating characteristic curve (TD AUCs) of DLP_INTG exhibited superior performance in comparison to the premise score in the training set ($p = 0.004$), the internal test set ($p = 0.046$), and the external test set ($p = 0.007$). The training set for DLP_INTG was created using the training and validation sets from DLP_DWI.

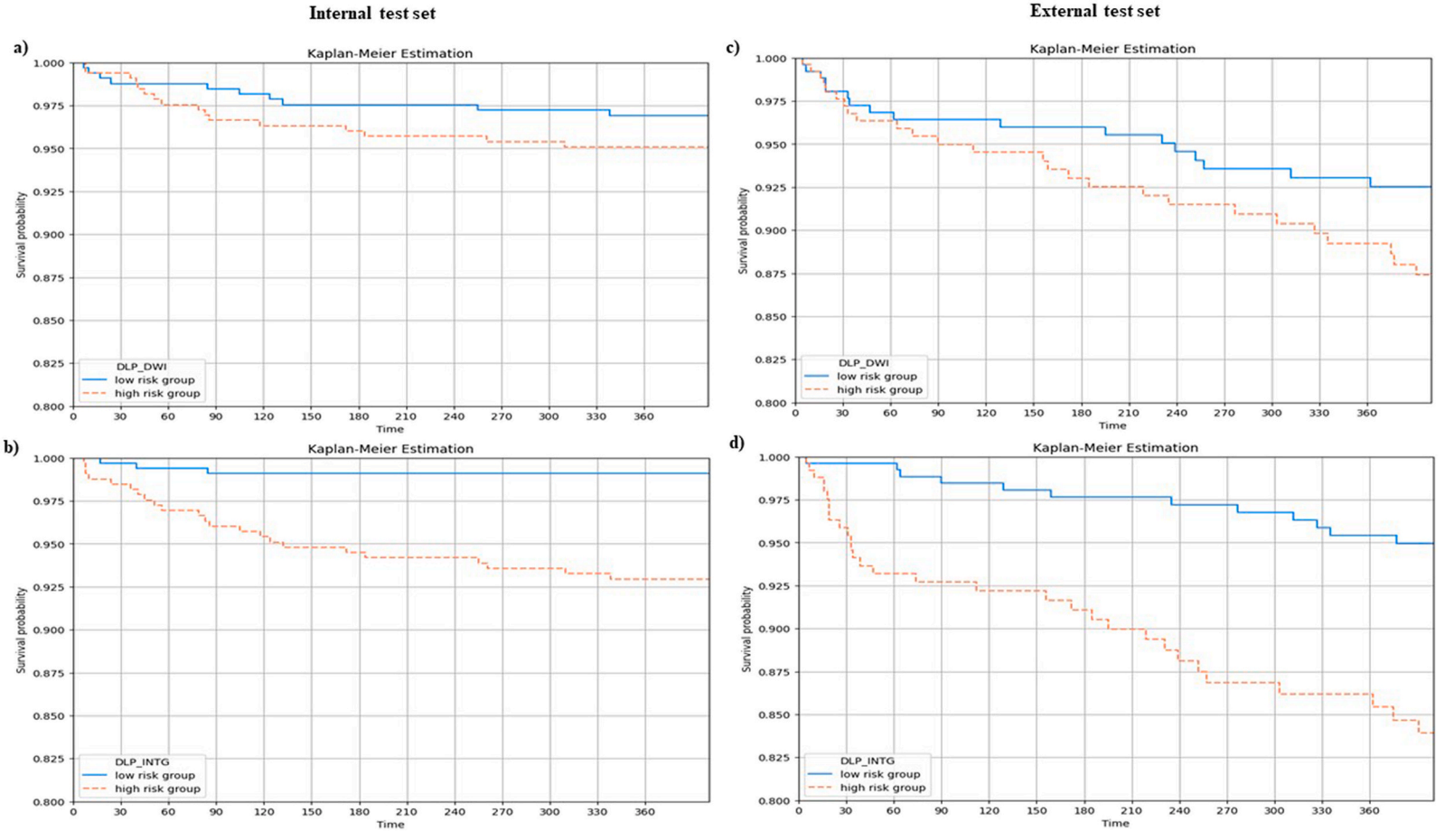


Fig. 3. Kaplan-Meier plot of deep learning mortality prediction model using diffusion weighted imaging (DLP_DWI) and integrated model (DLP_INTG). a), b) internal test set, and c), d) external test set, low vs. high-risk group, $p < 0.001$.

and ischemic heart disease compared to patients in the developmental cohort. Additionally, they presented with higher initial NIHSS scores, systolic blood pressure, and blood glucose levels upon admission (Supplementary Table S1).

3.3. Results of the prediction model

3.3.1. Performance of DLP_DWI

The performance of Densenet169 (AUC, 0.646) was slightly superior to that of ResNet-50 (AUC, 0.622), as shown in Supplementary Table S2. Consequently, we chose to use DenseNet-169 as deep learning mortality prediction model using brain DWI and ADC, referred to as DLP_DWI. Based on the Densenet169, the time-dependent area under the curve (TD AUC) of DLP_DWI was 0.643 [95 % confidence interval, CI: 0.524, 0.761], and the C-index of DLP_DWI was 0.765 [95% CI; 0.704, 0.826] in the internal test set. In the external test set, the TD AUC was 0.785 [95 % CI: 0.729, 0.842], and the C-index was 0.741 [95% CI: 0.717, 0.765] (Table 2). It achieved superior performance compared to infarction volume (TD AUC, 0.668 [95 % CI: 0.594, 0.742], $p = 0.018$) and the specific location of the ischemic lesion, particularly in cases of posterior circulation stroke (TD AUC, 0.517 [95% CI: 0.459, 0.582], $p < 0.001$) in the external test set. DLP_DWI was also an important predictor in the Cox regression model (Supplementary Table S3). The use of DLP_DWI was effective in discriminating the high-risk group of mortality, as shown in the Kaplan-Meier curve (Fig. 3a and 3c). The ablation study demonstrates that the integration of DWI and ADC significantly enhances model performance, achieving superior predictive accuracy (C-index: 0.642 and AUC: 0.646 internally; C-index: 0.731 and AUC: 0.769 externally) compared to the use of ADC or DWI alone. Notably, ADC's contribution marginally outweighs that of DWI, as evidenced by the performance metrics (Supplementary Table S4a).

3.3.2. Model selection and performance of the integrated model (DLP_INTG)

The variables encompassing age, NIHSS scores at admission, smoking history, atrial fibrillation, previous stroke, hypertension, diabetes, glucose levels, blood urea nitrogen (BUN), and creatinine were inputted into the integrated model. The results from the Cox regression model were shown in detail (Supplementary Table S3). Various machine learning model integrating DLP_DWI with clinical features (integrated model) outperforms models relying solely on clinical data (clinical model) or DLP_DWI, underscoring the value of multimodal data integration for optimal predictive performance (Supplementary Table S4b and S4c). The integrated model constructed with the Cox regression model exhibited superior performance in terms of the time-dependent area under the curve (TD AUC). It achieved TD AUC at 0.859 and C-index at 0.852, slightly surpassing the performance of the random forest classifier, which had TD AUC at 0.795 and C-index at 0.785, even though Catboost based classifier demonstrated the most robust performance among the diverse machine learning algorithms assessed on the internal test dataset (Supplementary Table S4). As a result, we selected the integrated Cox regression model as the final integrated model, DLP_INTG.

Moreover, we compared the traditional mortality score, premise score, and our integrated prediction model, DLP_INTG. In the training set, DLP_INTG achieved a TD AUC of 0.955 [95 % CI: 0.925, 0.984], outperforming the premise score, which had a TD AUC of 0.747 [95 % CI: 0.680, 0.815, $p = 0.004$]. In the internal test set, DLP_INTG also performed better with a TD AUC of 0.859 [95 % CI: 0.777, 0.941], compared to the premise score's TD AUC of 0.746 [95 % CI: 0.683, 0.813, $p = 0.046$], as shown in Table 2 and Fig. 4a. In

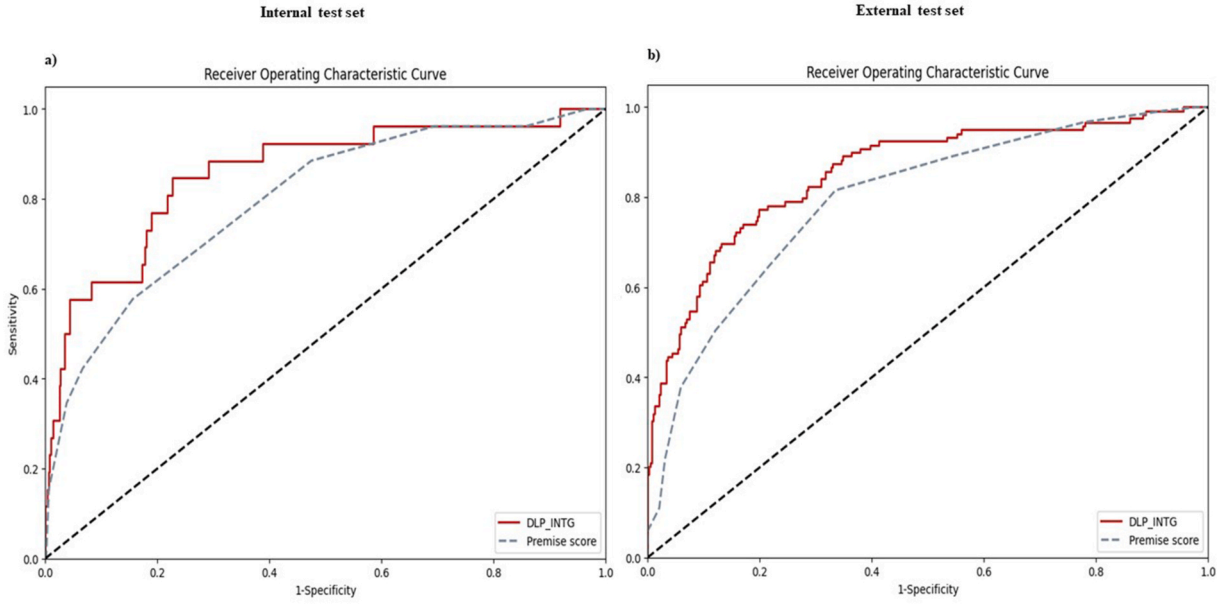


Fig. 4. Time dependent Area Under Receiver Operating Characteristic Curve (AUC) of integrated model (DLP_INTG) and premise score. a) internal test set, and b) external test set.

the external test set, the TD AUC of DLP_INTG, 0.876 [95 % CI: 0.828, 0.925] was significantly higher than that of the premise score, 0.808 [95 % CI: 0.756, 0.719, p = 0.007], as demonstrated in Table 2 and Fig. 4b. Similarly, the C-index of DLP_INTG was superior to that of the premise score in the external test set (0.842 vs. 0.780, p = 0.007), as indicated in Table 2. Furthermore, the use of DLP_INTG demonstrated effective discrimination of the high-risk group for mortality, as illustrated in the Kaplan-Meier curve (Fig. 3b and 3d). The performance of both DLP_DWI and DLP_INTG exhibited high specificity and negative predictive values, with additional details provided in Supplementary Table S5.

3.3.3. Explainability of DLP_DWI

The Grad-CAM analysis of DLP_DWI was assessed according to their volume and anatomical localization in true positive and true negative cases [24]. For anterior circulation, the Grad-CAM visualizations for true positive cases suggest a potential trend of focusing on lesions that extend to subcortical regions, as shown in Fig. 5(a–f).

4. Discussion

We developed the deep learning-based prediction model using brain DWI and ADC to capture the time dependent changes in imaging parameters for predicting one-year mortality after ischemic stroke, DLP_DWI, which showed the higher performance compared to the conventional imaging marker, such as infarction volume or posterior circulation. We also developed the integrated model, which incorporating both imaging and relevant clinical factors. DLP_INTG demonstrated superior performance compared to conventional mortality score. The survival curve demonstrated that the use of DLP_DWI and DLP_INTG effectively stratified higher-risk patients for one-year mortality, even during the acute stroke period. Furthermore, we conducted a validation of the prediction models,

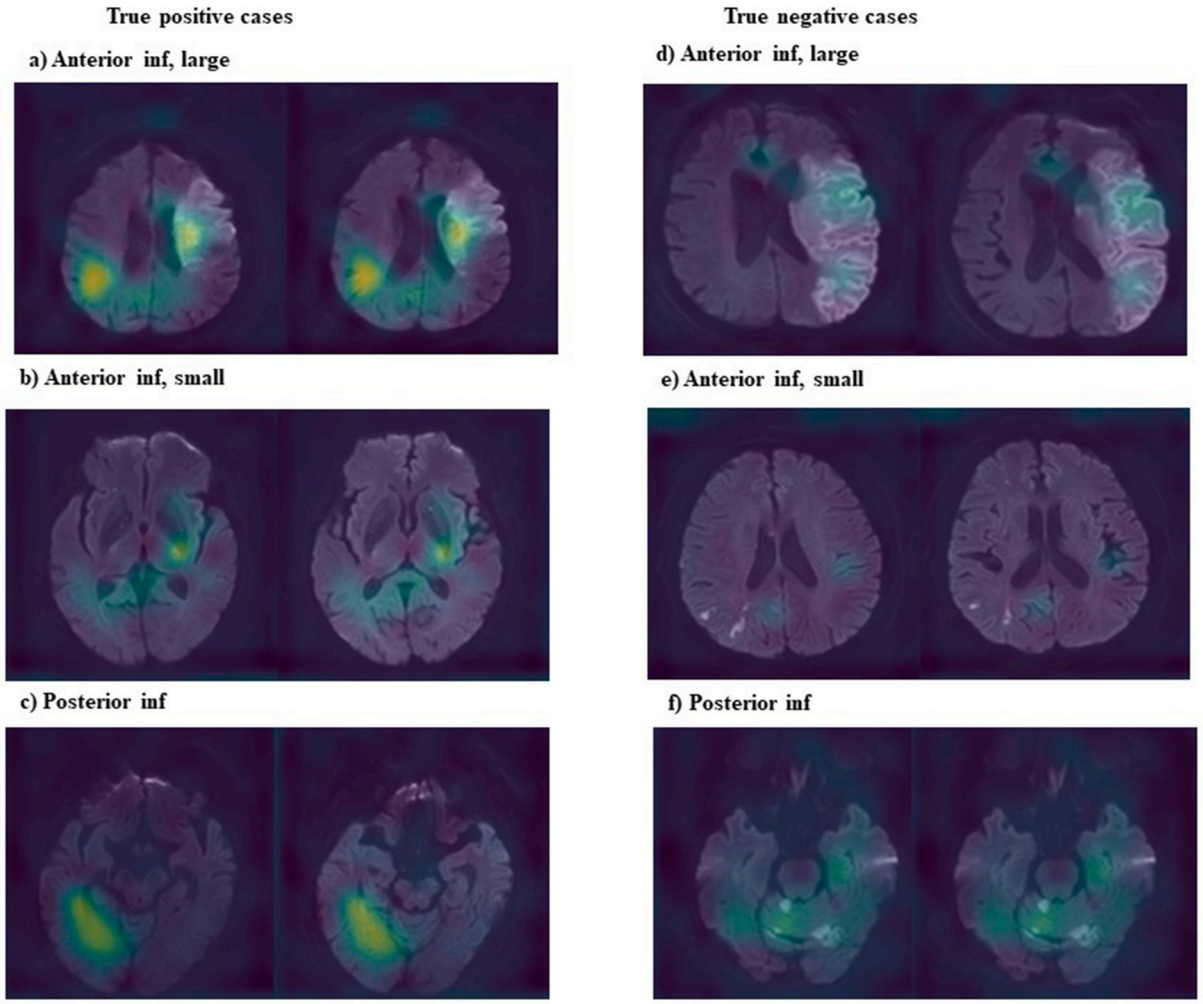


Fig. 5. Gradient-class activation maps (Grad-CAM). Panels a), b), and c) show true positive cases, while panels d), e), and f) illustrate true negative cases. Each set demonstrates Grad-CAM visualizations for different circulatory regions: anterior and large for a) and d), anterior and small for panels b) and e), and posterior circulation for panels c) and f), respectively.

DLP_DWI and DLP_INTG, using an external test cohort from a distinct geographical region and different medical center capacities, aiming to assess their reliability and generalizability.

The premise score comprises important clinical factors such as age and NIHSS score, and itemized imaging variables like the existence of posterior circulation and non-lacunar stroke.

We demonstrated that our DLP_DWI, which utilizes imaging data, showed performance comparable to that of the premise score, despite relying on a single feature. Moreover, our integrated model, DLP_INTG, utilizing imaging data and clinical variables, showed superior performance compared to the premise score. This finding supports the notion that radiological information derived from DLP_DWI might be more comprehensive, easily applicable in real-world practice.

The interest in the clinical application of CNNs for acute stroke imaging has been increasing, and has achieved state of-the-art performance in lesion segmentation, classification and so on [25,26]. However, the CNN based prediction models for outcome after stroke utilizing brain imaging exist, their primary emphasis has generally been on short-term outcomes following hyperacute stroke treatment, or the utilization of brain computer tomography (CT) imaging [11,27]. This differed from the scope of our study, which is primarily concerned with long-term mortality prediction after ischemic stroke using brain DWI and ADC, which was more precise methods to diagnose the ischemic stroke and reflect the characteristic of ischemic core, compared to brain CT.

Recently, Jin et al. reported on a nomogram for predicting 1-year mortality after ischemic stroke in patients admitted to the intensive care unit, utilizing data from the Medical Information Mart (MIMIC) database. Their nomogram proved helpful and easily applicable for predicting mortality and understanding the contribution of each variable. However, the patients in that study likely suffered severe strokes, necessitating neuro-intensive care interventions, such as invasive mechanical ventilation or mannitol injections, and did not incorporate radiologic information. These aspects significantly differ from the characteristics of the population and the scope of our study. We aimed to examine whether initial radiologic findings, specifically from DWI and ADC, at the time of stroke diagnosis, can add value in predicting long-term mortality in a time-dependent manner [28].

In addition, molten et al. reported that deep learning prediction model using only brain DWI to predict the functional outcome at 3-month after stroke was superior to conventional imaging markers, such as infarction volume [29]. In the same context with that study, we involved the development of a prototype prediction model that offers quantitative values derived from challenging-to-objectify images through the application of a deep learning model using brain DWI and ADC. However, the current study stood out in its uniqueness, as it implemented a survival-specified loss function that showed an effective performance in the time dependent prediction of mortality outcomes from brain DWI, ADC which has not been explored in stroke imaging research [4,30]. Further, this study extracted significant features from time-dependent contribution of brain imaging for predicting mortality and translated them into numerical values that are associated with the prognosis of stroke patients. This approach has the potential to streamline the process of collecting imaging data, reducing complexity, and facilitating the integration of clinical variables. In addition, our integrated model, DLP_INTG was developed based on the Cox regression model with merging clinical factors and outputs from DLP_DWI. We also used easily accessible and relevant clinical factors. This hybrid approach could combine the strengths of deep learning and traditional methodologies, leveraging deep learning's computational power for pattern recognition and the interpretive qualities of traditional methods [26,31]. Such an approach might enhance analytical performance and provide valuable clinical insights. In addition, we sought to enhance the explainability of our prediction model, particularly DLP_DWI, by employing Grad-CAM [15]. The attention map from Grad-CAM revealed a focus on the subcortical lesion, demonstrating a strong correlation with the corticospinal tract suggesting a link between severe, persistent motor deficits and an increased risk of morbidity and mortality [32]. Furthermore, subcortical lesions were significantly associated with a higher risk of developing post-stroke dementia, further influencing mortality rates. These findings are consistent with previous studies, underscoring the clinical importance of our prediction model. However, this observation does not establish direct causality. We emphasize the importance of interpreting Grad-CAM visualizations with caution, anchoring our interpretations in medical evidence, and adopting an evidence-based approach [33,34].

However, the limitations of relying solely on DLP_DWI for predictions exists. To address this, we developed and validated an integrated model, DLP_INTG, which complement radiologic and clinical data for improved accuracy. Future research with a larger dataset should explore these findings further to enhance the model's robustness and applicability [24]. These aspects should be investigated with a larger dataset in future studies. Our study had several limitations. Both developmental and external set were retrospectively collected, it is difficult to completely eliminate the potential for selection bias. In addition, the development and external datasets differed significantly due to their collection from hospitals specializing in different areas; the former was gathered from a general tertiary hospital, and the latter from a cardiovascular center specializing in high-risk cardiovascular surgeries. This discrepancy resulted in notable variations in patient characteristics, which were further influenced by geographical differences affecting access to medical care. Despite these challenges, our model outperformed traditional scoring methods, demonstrating its reliability and potential for broader application. However, the careful interpretation would be needed. Furthermore, instances of mortality constituted only 3.9 % of the dataset, presenting a significant imbalance issue common in clinical event analysis. The limited number of deceased patients in the developmental cohort and may have influenced the performance of our prediction model. However, we chose to use the original dataset and validated it in geographically distinct and differently-capacitated medical center to avoid the potential for overfitting or distortion of the dataset that can occur with oversampling techniques [35]. Recent technical advancements, such as the application of federated learning to imbalanced data, will be examined in future studies [36,37]. Additionally, this study was conducted on an Asian population. To address these limitations and enhance the generalizability of our findings, further validation through larger-scale and multicenter studies composing of various races with prospective design is warranted in the future.

We developed a deep learning model to predict survival rates following an ischemic stroke, incorporating both radiologic and clinical data. While we acknowledge that our research might not yet be ready for immediate clinical application, we believe in the importance of moving forward with more robust predictive model. These models could have a significant impact patient prognosis and

potentially reduce the substantial healthcare costs associated with ischemic stroke.

Sources of funding

This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea (NRF), South Korea funded by the Ministry of Education (MOE) (2022RIS-005), and by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), South Korea, funded by the Ministry of Health & Welfare, South Korea (grant number: HI22C1498).

Data availability statement

The data from this study are available upon request from the corresponding author. This research utilized datasets from 'The Open AI Dataset Project (AI-Hub, South Korea)'. All dataset information can be accessed through 'AI-Hub' (www.aihub.or.kr). Access to the dataset via AI-Hub requires IRB approval and is exclusively available to South Korean citizens. Proposals for collaboration will be evaluated for alignment with our criteria, and if deemed suitable, access to the Open AI Hub datasets will be facilitated.

CRediT authorship contribution statement

Changi Kim: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Joon-myoung Kwon:** Resources, Data curation. **Jiyeong Lee:** Resources. **Hongju Jo:** Resources. **Dowan Gwon:** Resources. **Jae Hoon Jang:** Resources. **Min Kyu Sung:** Resources. **Sang Won Park:** Supervision, Project administration, Funding acquisition. **Chulho Kim:** Writing – review & editing, Validation, Supervision, Funding acquisition. **Mi-Young Oh:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e31000.

References

[1] P.B. Gorelick, The global burden of stroke: persistent and disabling, Lancet Neurol. 18 (2019) 417–418.
[2] A.K. Boehme, C. Esenwa, M.S. Elkind, Stroke risk factors, genetics, and prevention, Circ. Res. 120 (2017) 472–495.
[3] S.H. Mittal, D. Goel, Mortality in ischemic stroke score: a predictive score of mortality for acute ischemic stroke, Brain Circul. 3 (2017) 29.
[4] T. Gattringer, A. Posekany, K. Niederkorn, M. Knoflach, B. Poltrum, S. Mutzenbach, et al., Predicting early mortality of acute ischemic stroke: score-based approach, Stroke 50 (2019) 349–356.
[5] G. Saposnik, A.K. Guzik, M. Reeves, B. Ovbiagele, S.C. Johnston, Stroke prognostication using age and nih stroke scale: span-100, Neurology 80 (2013) 21–28.
[6] A.C. Flint, B.S. Faigeles, S.P. Cullen, H. Kamel, V.A. Rao, R. Gupta, et al., Thrive score predicts ischemic stroke outcomes and thrombolytic hemorrhage risk in vista, Stroke 44 (2013) 3365–3369.
[7] S. Rangaraju, C. Streib, A. Aghaebrahim, A. Jadhav, M. Frankel, T.G. Jovin, Relationship between lesion topology and clinical outcome in anterior circulation large vessel occlusions, Stroke 46 (2015) 1787–1792.
[8] S.F. Zaidi, A. Aghaebrahim, X. Urra, M.A. Jumaa, B. Jankowitz, M. Hammer, et al., Final infarct volume is a stronger predictor of outcome than recanalization in patients with proximal middle cerebral artery occlusion treated with endovascular therapy, Stroke 43 (2012) 3238–3244.
[9] E. Adeli, Q. Zhao, N.M. Zahr, A. Goldstone, A. Pfefferbaum, E.V. Sullivan, et al., Deep learning identifies morphological determinants of sex differences in the pre-adolescent brain, Neuroimage 223 (2020) 117293.
[10] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[11] Y. Yu, S. Christensen, J. Ouyang, F. Scalzo, D.S. Liebeskind, M.G. Lansberg, et al., Predicting hypoperfusion lesion and target mismatch in stroke from diffusion-weighted mri using deep learning, Radiology 307 (2022) e220882.
[12] S. Bacchi, T. Zerner, L. Oakden-Rayner, T. Kleinig, S. Patel, J. Jannes, Deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes: a pilot study, Acad. Radiol. 27 (2020) e19–e23.
[13] M.F. Gensheimer, B. Narasimhan, A scalable discrete-time survival model for neural networks, PeerJ 7 (2019) e6257.
[14] M. Talo, O. Yildirim, U.B. Baloglu, G. Aydin, U.R. Acharya, Convolutional neural networks for multi-class brain disease detection using mri images, Comput. Med. Imag. Graph. 78 (2019) 101673.

[15] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.
[16] J.R. Quinlan, Induction of decision trees, Mach. Learn. 1 (1986) 81–106.
[17] C. Bentéjac, A. Csörgő, G. Martínez-Muñoz, A comparative analysis of gradient boosting algorithms, Artif. Intell. Rev. 54 (2021) 1937–1967.
[18] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin, Catboost: unbiased boosting with categorical features, Adv. Neural Inf. Process. Syst. (2018) 31.
[19] P. Blanche, J.F. Dartigues, H. Jacqmin-Gadda, Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks, Stat. Med. 32 (2013) 5381–5397.
[20] E. Longato, M. Vettoretti, B. Di Camillo, A practical perspective on the concordance index for the evaluation and selection of prognostic time-to-event models, J. Biomed. Inf. 108 (2020) 103496.
[21] W.J. Youden, Index for rating diagnostic tests, Cancer 3 (1950) 32–35.
[22] J.M. Bland, D.G. Altman, Survival probabilities (the kaplan-meier method), BMJ 317 (1998) 1572–1580.
[23] G.S. Collins, J.B. Reitsma, D.G. Altman, K.G. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod) the tripod statement, Circulation 131 (2015) 211–219.
[24] E. Zurcher, B. Richoz, M. Faouzi, P. Michel, Differences in ischemic anterior and posterior circulation strokes: a clinico-radiological and outcome analysis, J. Stroke Cerebrovasc. Dis. 28 (2019) 710–718.
[25] A.D. Yao, D.L. Cheng, I. Pan, F. Kitamura, Deep learning in neuroradiology: a systematic review of current algorithms and approaches for the new wave of imaging technology, Radiol.: Artif. Intell. 2 (2020) e190026.
[26] L. Cui, H. Li, W. Hui, S. Chen, L. Yang, Y. Kang, et al., A deep learning-based framework for lung cancer survival analysis with biomarker interpretation, BMC Bioinf. 21 (2020) 1–14.
[27] A. Nielsen, M.B. Hansen, A. Tietze, K. Mouridsen, Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning, Stroke 49 (2018) 1394–1401.
[28] G. Jin, W. Hu, L. Zeng, B. Ma, M. Zhou, Prediction of long-term mortality in patients with ischemic stroke based on clinical characteristics on the first day of icu admission: an easy-to-use nomogram, Front. Neurol. 14 (2023) 1148185.
[29] E. Moulton, R. Valabregue, M. Piotin, G. Marnat, S. Saleme, B. Lapergue, et al., Interpretable deep learning for the prognosis of long-term functional outcome post-stroke using acute diffusion weighted imaging, J. Cerebr. Blood Flow Metabol. 43 (2023) 198–209.
[30] H. Kim, J.M. Goo, K.H. Lee, Y.T. Kim, C.M. Park, Preoperative ct-based deep learning model for predicting disease-free survival in patients with lung adenocarcinomas, Radiology 296 (2020) 216–224.
[31] J.G. Nam, H.-R. Kang, S.M. Lee, H. Kim, C. Rhee, J.M. Goo, et al., Deep learning prediction of survival in patients with chronic obstructive pulmonary disease using chest radiographs, Radiology 305 (2022) 199–208.
[32] J. Liu, C. Wang, W. Qin, H. Ding, J. Guo, T. Han, et al., Corticospinal fibers with different origins impact motor outcome and brain after subcortical stroke, Stroke 51 (2020) 2170–2178.
[33] M. Soltanisarvestani, N. Lynskey, S. Gray, J.M. Gill, J.P. Pell, N. Sattar, et al., Associations of grip strength and walking pace with mortality in stroke survivors: a prospective study from UK biobank, Scand. J. Med. Sci. Sports (2023).
[34] L.M. Allan, E.N. Rowan, M.J. Firbank, A.J. Thomas, S.W. Parry, T.M. Polvikoski, et al., Long term incidence of dementia, predictors of mortality and pathological diagnosis in older stroke survivors, Brain 134 (2011) 3716–3727.
[35] M. Nakamura, Y. Kajiwara, A. Otsuka, H. Kimura, Lvq-smote - learning vector quantization based synthetic minority over-sampling technique for biomedical data, BioData Min. 6 (2013) 16.
[36] M.J. Sheller, B. Edwards, G.A. Reina, J. Martin, S. Pati, A. Kotrotsou, et al., Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data, Sci. Rep. 10 (2020) 12598.
[37] L. Huang, A.L. Shea, H. Qian, A. Masurkar, H. Deng, D. Liu, Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records, J. Biomed. Inf. 99 (2019) 103291.