P-9

Long-term Survival Prediction in Stroke and Traumatic Brain Injury using Machine Learning Algorithms

국민건강보험 일산병원 National Health Insurance Service Ilsan Hospital

Nayoon Yoo[1], Dougho Park[2], Jong Hun Kim[3], Hyoung Seop Kim[1]

[1]Department of Physical Medicine and Rehabilitation, National Health Insurance Service Ilsan Hospital
[2]Department of Rehabilitation Medicine, Pohang Stroke and Spine Hospital
[3]Department of Neurology, National Health Insurance Service Ilsan Hospital

Objective

Long-term prognosis after stroke or traumatic brain injury can be elucidated by machine learning algorithms recently. We propose machine learning (ML) models that predict long-term survival after stroke and TBI based on a combined national health insurance (NHIS) and its self-run hospital database.

Methods

This retrospective cohort study included adults aged ≥20 years who were diagnosed with stroke or TBI between 2009 and 2018. We categorized the participants into long-term care insurance (LTCI) and non-LTCI beneficiaries. The dependent variable was survival until December 31, 2018. Random forest (RF), extreme gradient boosting (XGB), and algorithm stacking methods were used for prediction.

원내자료		공단자료	노인장기요양보험 자료
연령	전체 입원기간	장애 유형	장기요양서비스 종류
성별	일반병동 입원기간	장애 등급	인증 점수
진단명	협진 횟수	가입자 유형	
생년월일	기관절개술 여부	보험료 분위	
사망여부	경관급식 여부		
신체계측	합병증		
혈액검사	영상의학 판독 자료		

Results

Of the total patients, 1839 patients were in the LTCI group, while 2950 patients were in the non-LTCI group. The LTCI group was older (76.1 ± 9.4 years old vs. 62.4 ± 13.6 years old; p <0.001) and had a lower survival rate (52.3% vs. 23.6%; p <0.001) (Table 1). In the LTCI group, the stacking algorithms showed the highest area under the Receiver Operating Characteristic curve (AUC) value of 0.86. Both RF and XGB had an AUC of 0.83 (Table 2). In the non-LTCI group, XGB showed the highest AUC value at 0.81, and the RF and stacking algorithms showed AUCs of 0.78 and 0.8, respectively (Table 3).

Table 1. Baseline Characteristics of Entire Patients.

Variables		LTCI group (n = 1839)	Non-LTCI group (n = 2950)	p-value
Death, n (%)		962 (52.3)	695 (23.6)	<0.001
Survival length, days		1668.1 ± 1135.4	1831.7 ± 1114.6	<0.001
Age, years		76.1 ± 9.4	62.4 ± 13.6	<0.001
Male, n (%)		808 (43.9)	1813 (61.5)	<0.001
Hospital days		25.0 ± 40.6	16.1 ± 24.6	<0.001
Days in the intensive care unit		9.7 ± 35.8	5.9 ± 16.9	<0.001
Consultation claims, n (%)		9.8 ± 9.9	7.1 ± 8.3	<0.001
Tube feeding, n (%)		668 (36.3)	405 (13.7)	<0.001
Hospitalized pneumonia, n (%)		1107 (60.2)	1176 (39.9)	<0.001
Tracheostomy, n (%)		61 (3.3)	52 (1.8)	0.001
Pressure ulcer, n (%)		113 (6.1)	62 (2.1)	<0.001
Smoking	Current	954 (51.9)	1249 (42.3)	<0.001
	Quitter	718 (39.0)	1048 (35.5)	
	Non-smoker	167 (9.1)	653 (22.1)	
Alcohol, n (%)	Current	1322 (71.9)	1604 (54.4)	<0.001
	Quitter	297 (16.2)	409 (13.9)	
	Non-drinker	220 (12.0)	937 (31.8)	
Albumin, g/dL		3.7 ± 0.5	4.0 ± 0.5	<0.001
Blood urea nitrogen, mg/dL		19.3 ± 12.4	16.7 ± 10.3	<0.001
Creatinine, mg/dL		1.2 ± 1.3	1.1 ± 1.4	0.014
Glucose Ante Cibum, mg/dL		133.8 ± 54.3	129.4 ± 49.3	0.004
Hemoglobin, g/dL		13.0 ± 2.0	13.8 ± 2.0	<0.001
HbA1c, %		6.6 ± 1.4	6.5 ± 1.3	0.032
Cerebral infarction, n (%)		1405 (76.4)	1571 (53.3)	<0.001
Subarachnoid hemorrhage, n (%)		134 (7.3)	337 (11.4)	<0.001
Subdural hemorrhage, n (%)		182 (9.9)	227 (7.7)	0.009
Intracranial hemorrhage, n (%)		474 (25.8)	637 (21.6)	0.001
Intraventricular hemorrhage, n (%)		144 (7.8)	172 (5.8)	0.008
Traumatic brain injury, n (%)		85 (4.6)	102 (3.5)	0.052

Abbreviation: LTCI, long-term care insurance.

Table 2. Machine Learning Performance in the LTCI group.

Algorithms	AUC	Sensitivity	Specificity	PPV	NPV
Random Forest	0.834	0.764	0.903	0.878	0.808
XGB	0.834	0.768	0.900	0.874	0.810
Stacking algorithms†	0.864	0.914	0.814	0.833	0.903

Abbreviation: AUC, area under the Receiver Operating Characteristic curve; NPV, negative predictive value; PPV, positive predictive value; XGB, extreme gradient boosting.
†stacked algorithms: neural network, support vector machines, k-nearest neighbors, classification and regression trees, random forest, gradient boosting machine, and extreme gradient boosting.

Table 3. Machine Learning Performance in the Non-LTCI group.

Algorithms	AUC	Sensitivity	Specificity	PPV	NPV
Random Forest	0.780	0.953	0.606	0.873	0.820
XGB	0.810	0.961	0.659	0.889	0.856
Stacking algorithms†	0.801	0.947	0.655	0.887	0.813

Abbreviation: AUC, area under the Receiver Operating Characteristic curve; NPV, negative predictive value; PPV, positive predictive value; XGB, extreme gradient boosting.
†stacked algorithms: neural network, support vector machines, k-nearest neighbors, classification and regression trees, random forest, gradient boosting machine, and extreme gradient boosting.

Conclusion

ML algorithms applied in this study showed valid and high accuracy and AUC values for predicting long-term survival in patients with stroke and TBI. This study is relevant because we could utilize various variables by combining the NHIS and its self-run hospital.