Literature Review

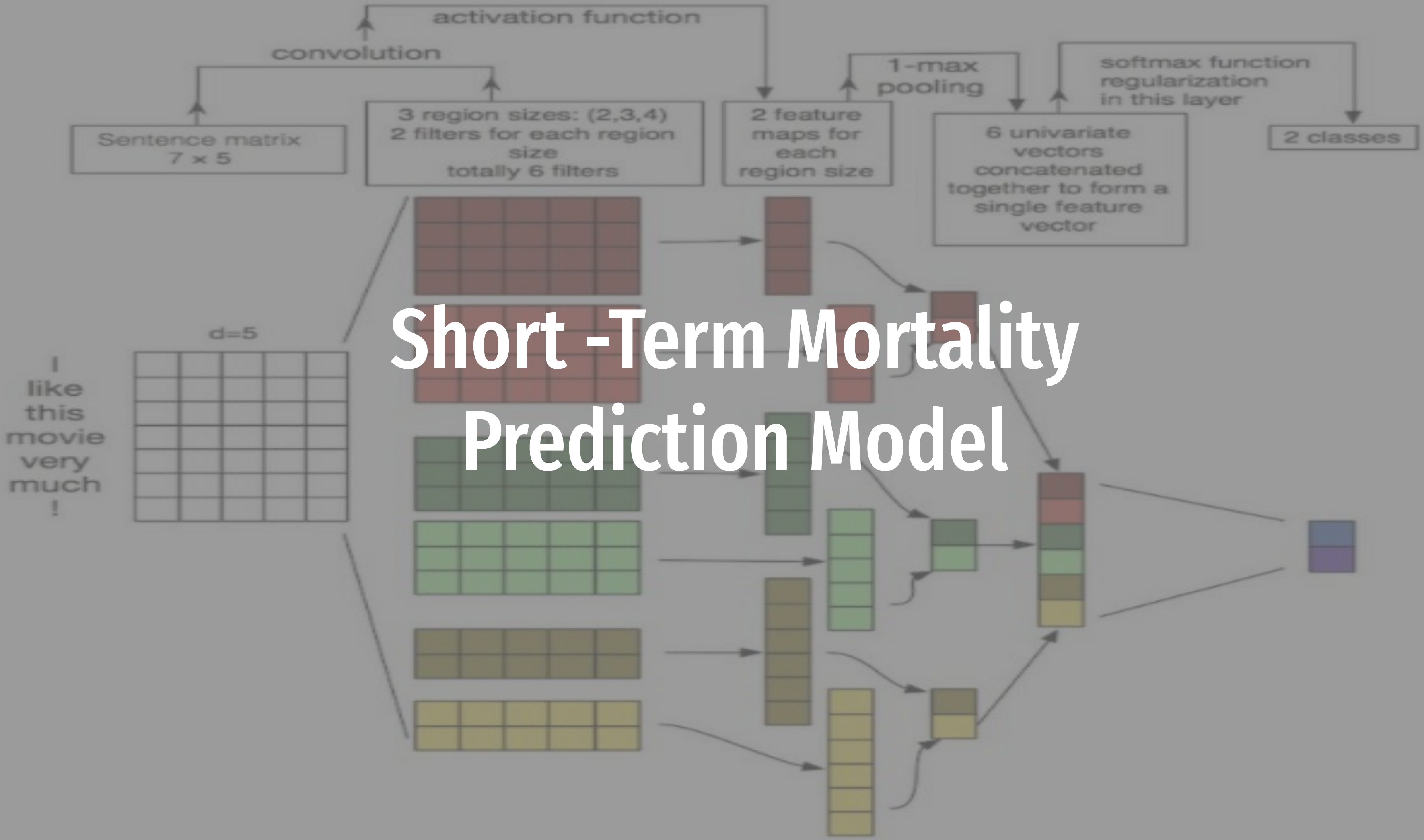
Short term mortality prediction model analysis
&
comprehend ensemble model

21.Oct.15(Fri)

암관리학과 권회준

Contents

Short -Term Mortality Prediction Model



Original Investigation

Time-Limited Trials of Intensive Care for Critically Ill Patients With Cancer
How Long Is Long Enough?

Mark G. Shrime, MD, MPH, PhD; Bart S. Ferket, MD, PhD; Daniel J. Scott, PhD; Joon Lee, PhD;
Diana Barragan-Bradford, MD; Tom Pollard, PhD; Yaseen M. Arabi, MD; Hasan M. Al-Dorzi, MD;
Rebecca M. Baron, MD; M. G. Myriam Hunink, MD, PhD; Leo A. Celi, MD, MS, MPH; Peggy S. Lai, MD, MPH

Topic : 중환자실 암 환자의 단기 사망률에 대한 최적의 집중 치료 기간을 식별

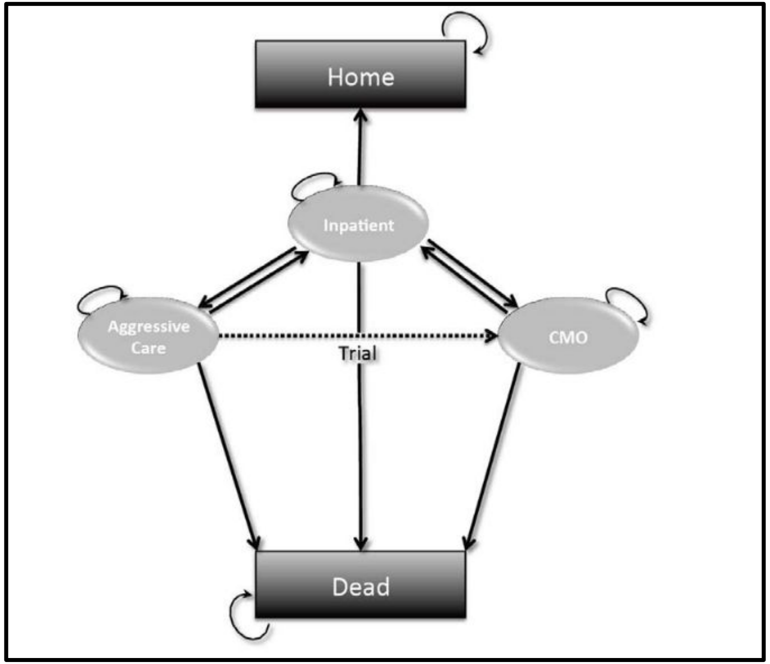
Study population : 2001-2007 입원한 위독한 암 환자 920명

Outcomes : 30-day all-cause mortality and mean survival duration.

Data Source : A public-access intensive care unit database (MIMIC-II)*

* Multiparameter Intelligent Monitoring in Intensive Care II

Short - Term Mortality Prediction Model/ Review



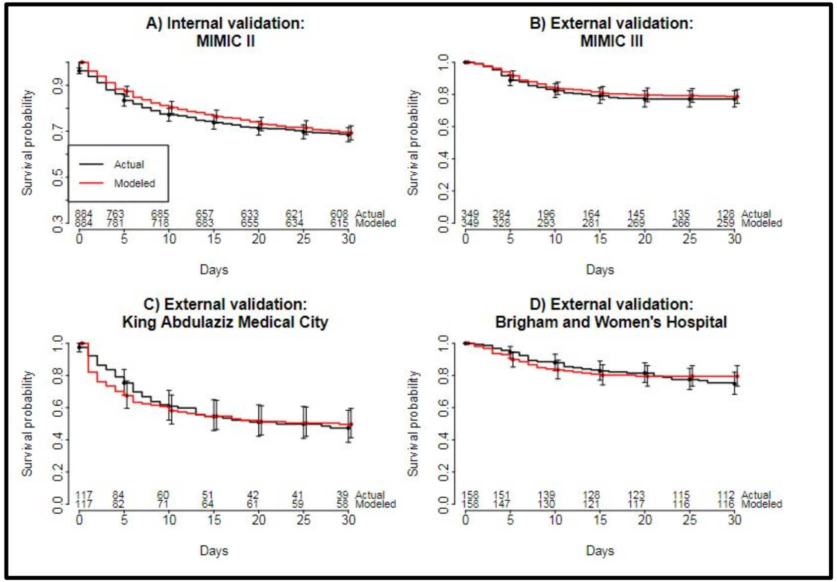
Model: State Transition Model(STM, Markov model)

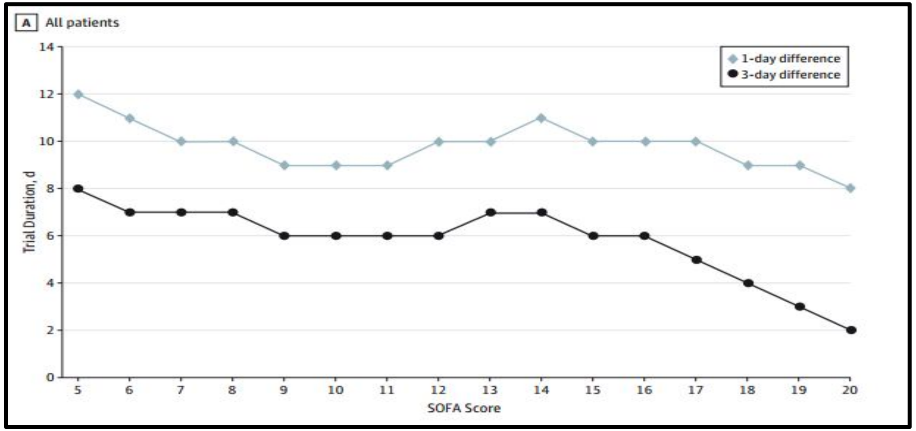
Parameter:
1. Daily probability of improving given CMO* status
2. Daily probability of discharge home from an inpatient floor
3. Daily probability of return to ICU from an inpatient floor

Variable: SOFA* SCORE

Software: R, Tree Age

* Comfort Measures Only; Sequential Organ Failure Assessment;





Validation: 1 internal, 3 external data

Point

외부 데이터를 사용하여 검증하였음에도 높은 예측률 보임
SOFA SCORE 변수가 가지는 영향성 확인

JAMA Network Open

Original Investigation | Oncology

Machine Learning Approaches to Predict 6-Month Mortality Among Patients With Cancer

Ravi B. Parikh, MD, MPP; Christopher Manz, MD; Corey Chivers, PhD; Susan Harkness Regli, PhD; Jennifer Braun, MHA; Michael E. Draugelis, MS; Lynn M. Schuchter, MD; Lawrence N. Shulman, MD; Amol S. Navathe, MD, PhD; Mitesh S. Patel, MD, MBA; Nina R. O'Connor, MD

Topic : 암 환자 사망률 예측을 위한 머신러닝 알고리즘 개발 및 검증

Study population : 16.2 - 16.7, 종양 또는 혈액 관련 암 진료를 받은 성인 26,525명

Outcomes : 180-day mortality from the index encounter;

Data Source : University of Pennsylvania Health System EHR

Table 2. Performance Metrics of Machine Learning Models[a]

Algorithm	Positive Predictive Value[b]	AUC[b]	Accuracy	Specificity
Random forest	0.513[c]	0.88[c]	0.96[c]	0.99[c]
Gradient boosting classifier	0.494	0.87	0.96[c]	0.99[c]
Logistic regression	0.447	0.86	0.95	0.99[c]

Abbreviation: AUC, area under the receiver operating characteristic curve.

[a] Positive predictive value, accuracy, and specificity were determined by setting the alert rate in the test set for each algorithm to 0.02. At this prespecified alert rate, the 6-month mortality risk threshold was 0.27 for the random forest model; 0.15 for the gradient boosting model; and 0.33 for the logistic regression model.

[b] Coprimary performance metric.

[c] Refers to the best-performing model(s) for each performance metric.

Model: Logistic regression, Gradient boosting, and Random forest algorithms.

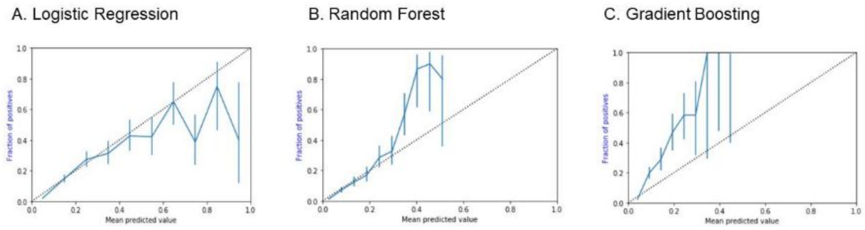
Selected Model: Gradient boosting.

Variable:
1. Demographic variables (eg, age and sex)
2. Elixhauser comorbidities(31)
3. laboratory(83) and select electrocardiogram data(9)

Train/Test: 18 567 (70.0%) / 7958 (30.0%)

Software: SAS, Python

eFigure 2. Model Calibration Plots



Footnote: Calibration plots describing observed (y-axis) vs. predicted (x-axis) mortality for the logistic regression (A), random forest (B), and gradient boosting (C) models. Each point represents one of ten bins of predicted probability of 180-day mortality for each model. The observed rate of 180-day mortality for each of the probability bins is plotted on the y-axis. The 45-degree dotted diagonal line represents points along a perfectly-calibrated model. The bars show 95% confidence intervals around the observed probability.

Point

과적합에도 불구하고, gradient boosting 및 random forest 모델은 우수한 판별력과 홀드아웃 유효성 검사 세트에서 우수한 PPV를 보임

실제 모델의 성능과 임상에서의 타당한 부분은 차이가 존재

JAMA Oncology | Original Investigation

Validation of a Machine Learning Algorithm to Predict 180-Day Mortality for Outpatients With Cancer

Christopher R. Manz, MD; Jinbo Chen, PhD; Manqing Liu, MHS; Corey Chivers, PhD; Susan Harkness Regli, PhD; Jennifer Braun, MHA; Michael Draugelis, MS; C. William Hanson, MD; Lawrence N. Shulman, MD; Lynn M. Schuchter, MD; Nina O'Connor, MD; Justin E. Bekelman, MD; Mitesh S. Patel, MD, MBA; Ravi B. Parikh, MD, MPP

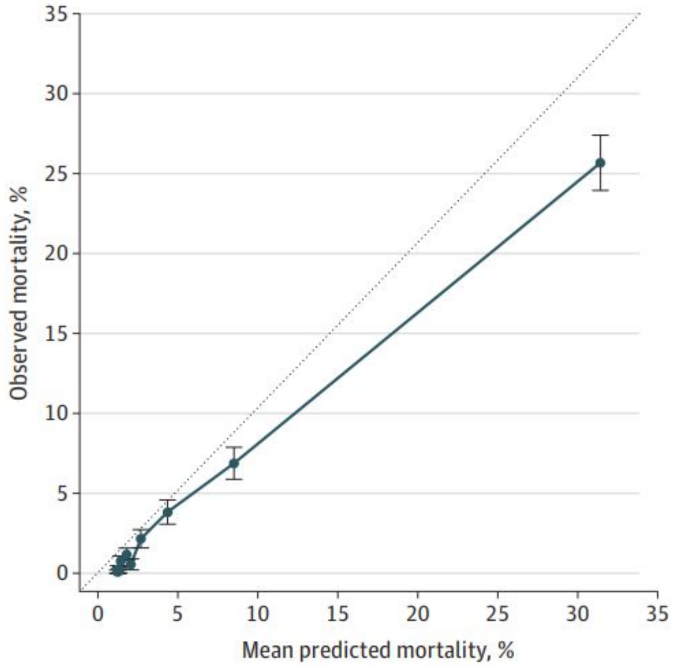
Topic : 180일 사망률 위험에 대해 실시간 예측을 진행하는 EHR 내장 ML 알고리즘을 검증

Study population : 19.3 - 19.4, 종양 진료를 받은 외래 환자 24,582명

Outcomes : 180-day mortality from the index encounter;
Performance metric was the area under the receiver operating characteristic curve;

Data Source : University of Pennsylvania Health System EHR

Figure 2. Algorithm Calibration



Model: Gradient boosting,

Variable:
Age, Sex, Ethics, Marital Status, Insurance, ECOG*, Perforamnce status, Elixhauser comorbidity socre, cancer state, pracitce site

Software: Python

* Eastern Cooperative Oncology Group;

Short - Term Mortality Prediction Model/ Review

Table 3. Comparison of ML Algorithm Against Standard Prognostic Indices in Oncology

	Prognostic index	
Variable	ECOG (n = 6225)	Elixhauser (n = 24 582)
Baseline classifier AUC (95% CI)	0.72 (0.70-0.75)	0.69 (0.68-0.71)
ΔAUC (95% CI) between ML algorithm and baseline classifier	0.17 (0.14-0.19)	0.20 (0.18-0.21)
Baseline classifier PPV, No.	0.27	0.10
ΔPPV between ML algorithm and baseline classifier, No.	0.18	0.36
NRI (95% CI) from enhanced vs baseline classifier	0.09 (0.04-0.14)	0.23 (0.20-0.27)

Abbreviations: AUC, area under the receiver operating characteristic curve; ECOG, Eastern Cooperative Oncology Group; ML, machine learning; NRI, net reclassification index; PPV, positive predictive value; ΔAUC, difference in AUC; ΔPPV, difference in PPV.

Point

기존 모델에 비해 예후 예측이 유리하게 비교됨을 통해
임상의들의 의사결정도구로서 유용하게 사용 가능성 내재

Original Investigation | Oncology

Development and Application of a Machine Learning Approach to Assess Short-term Mortality Risk Among Patients With Cancer Starting Chemotherapy

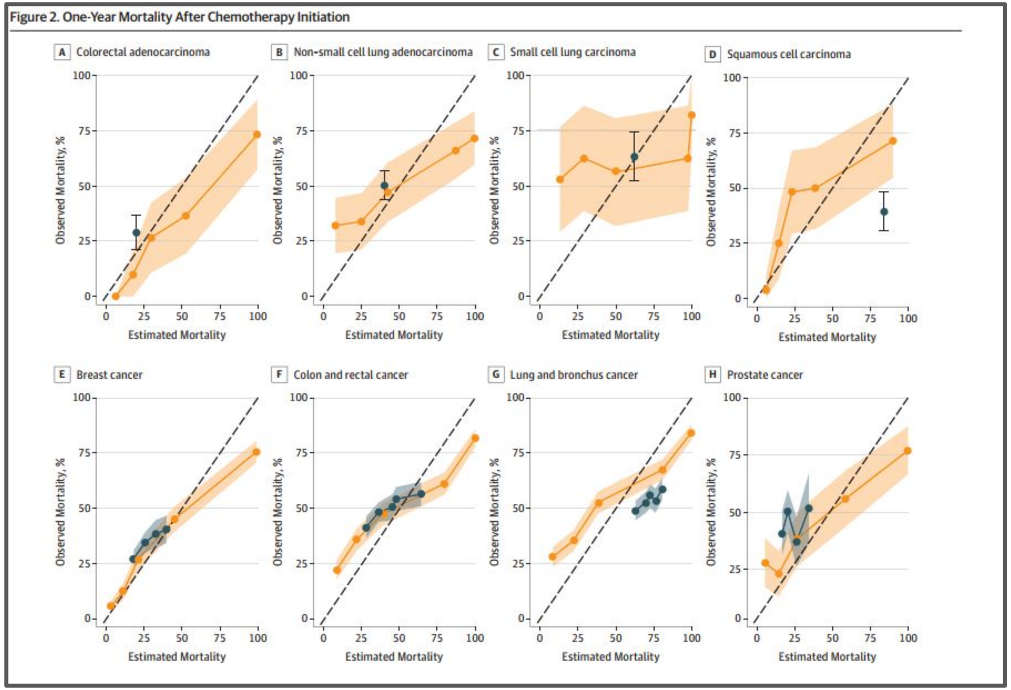
Aymen A. Elfiky, MD, MPH, MSc, MBA; Maximilian J. Pany, BA; Ravi B. Parikh, MD, MPP; Ziad Obermeyer, MD, MPhil

Topic : 치료 첫날 이전에 사용할 수 있는 데이터만 사용하여 사망률을 예측하는 모델 생성 및 검증

Study population : 2004-2011, 26,946명

Outcomes : 1. Thirty-day mortality from the first day of a new chemotherapy regimen;
2. Model discrimination by predicted mortality risk decile among patients receiving palliative chemotherapy and 180-day mortality from the first day of a new chemotherapy regimen;

Data Source : Dana-Farber/Brigham and Women’s Cancer Center



Model: gradient-boosted trees (Gradient boosting)

Variable:
1. patient demographics
2. prescribed medications
3. diagnoses and procedures
5. care utilization (e.g., inpatient, outpatient, and emergency encounters)
6. vital signs and laboratory results

Software: R

Short - Term Mortality Prediction Model / Review

Table 3. Selected Predictors by Risk Decile and Model Variance Explained

Predictor	Risk Decile			Model Variance Explained, %
	Top	Median	Bottom	
Cancer of the brain and with other nervous system areas, %	6.7	1.6	0.0	1.94
Demographics				
Mean age, y	62.3	62.1	51.9	1.30
Female, %	56.4	60.9	86.9	1.17
Black, %	3.8	3.6	3.4	0.07
Mean comorbidity score[a]	5.14	3.44	2.01	0.03
Prior diagnoses, %				
Ascites	0.31	0.07	0.01	0.39
Mouth disorder	0.02	0.01	0.01	0.25
Nausea and vomiting	0.23	0.09	0.01	0.18
Lower respiratory tract disorders	2.10	1.36	0.16	0.03
Secondary malignant neoplasm	5.57	1.69	0.36	0.01
Failure to thrive	0.05	0.01	0.00	0.01
Nutritional disorders	2.53	1.63	0.54	0.00
Malaise and fatigue	0.22	0.08	0.02	0.00
Medications, %				
Corticosteroids	0.53	0.00	0.00	0.15
Opioids	0.29	0.00	0.00	0.05
Anxiolytics	0.60	0.24	0.18	0.02
Cathartics	0.54	0.00	0.00	0.00
Vital signs[b]				
Maximum pulse, bpm (baseline)	106.1	95.7	87.1	0.37
Maximum weight, kg (baseline)	79.5	80.2	76.8	0.11
Weight, SD, kg (baseline)	3.1	2.1	1.3	0.06
Minimum pulse, bpm (recent)	83.4	72.8	63.0	0.01
Weight change, kg[c]	−3.1	−1.0	0.1	0.00
Laboratory findings[d]				
Maximum C-reactive protein level, mg/L	93.9	65.6	2.2	0.19
Maximum ALT level, U/L	75.9	57.3	24.3	0.07
Maximum AST level, U/L	73.7	54.2	23.8	0.09
Maximum white blood cell count, $\times 10^3$/μL	13.9	12.4	9.8	0.03
Maximum alkaline phosphatase, IU/L	199.5	128.7	76.5	0.02
Mean lymphocyte count, $\times 10^3$/μL	1.0	1.3	1.8	0.00
Mean platelet count, $\times 10^3$/μL	251.8	241.4	265.1	0.00
Ejection fraction, %	54.4	48.0	51.9	0.01
Total linear terms	NA	NA	NA	13.6
Total nonlinear terms and interactions	NA	NA	NA	86.4

Point

다종 암에 대해 정확한 식별 가능 , 화학요법 시작전 치료 계획에 대해 환자의 의사결정을 안내하는데 도움 줄 수 있음

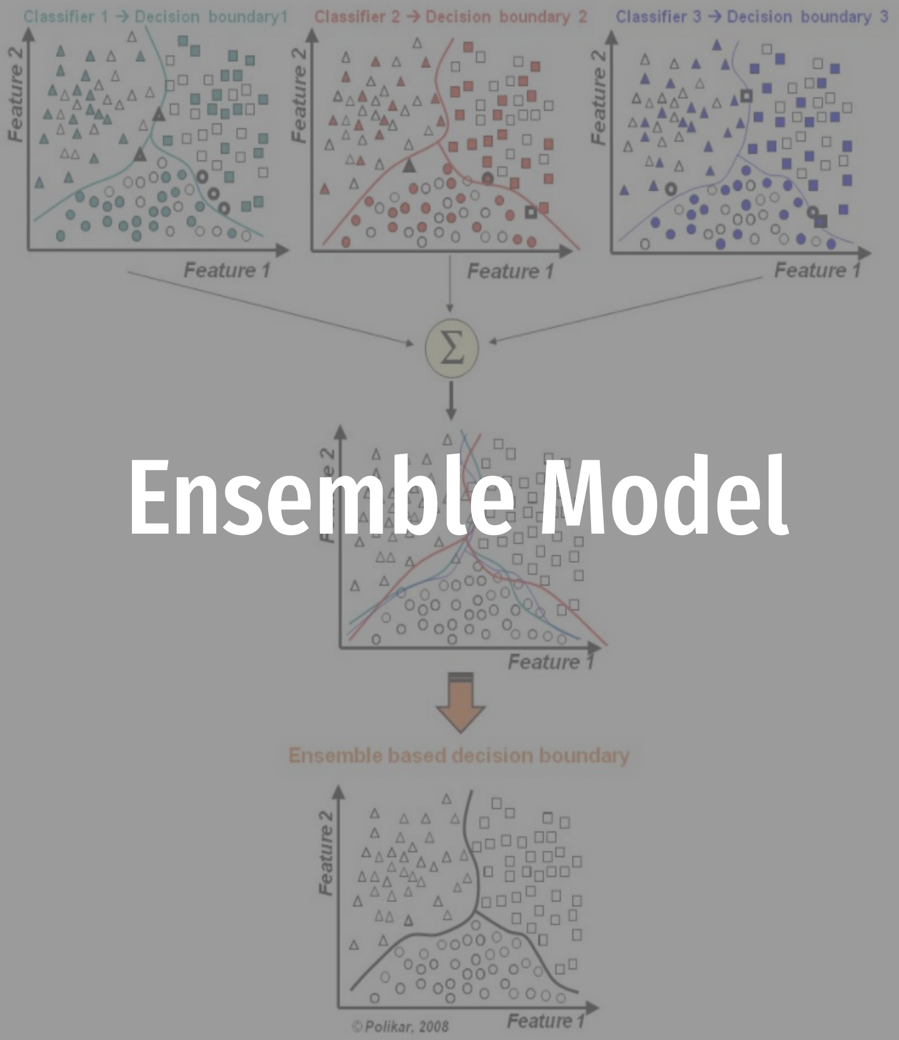
1. markov 모형을 제외한 모든 ML 모델에서 상당한 데이터를 요구 특히 다종암을 분류하려 mortality 예측 모델은 5000개 이상의 변수 사용;
→ 많은 데이터 필요

2. 모델을 하나로 확정 짓는 것이 아니라 모델 후보 선정하여 구성한 뒤 비교하여 최선의 모형 구성

3. 첫번째 논문에서 SOFA Score 사용 제한 문제; ICU 인 경우에만 해당

4. 두번째 세번째 논문에서 Elixhauser comorbidity score 차용, 국내는 Chalson comorbidity 을 주로 사용하는데 index의 차이가 성능에 얼마나 미치는지 파악 불가

Ensemble Model



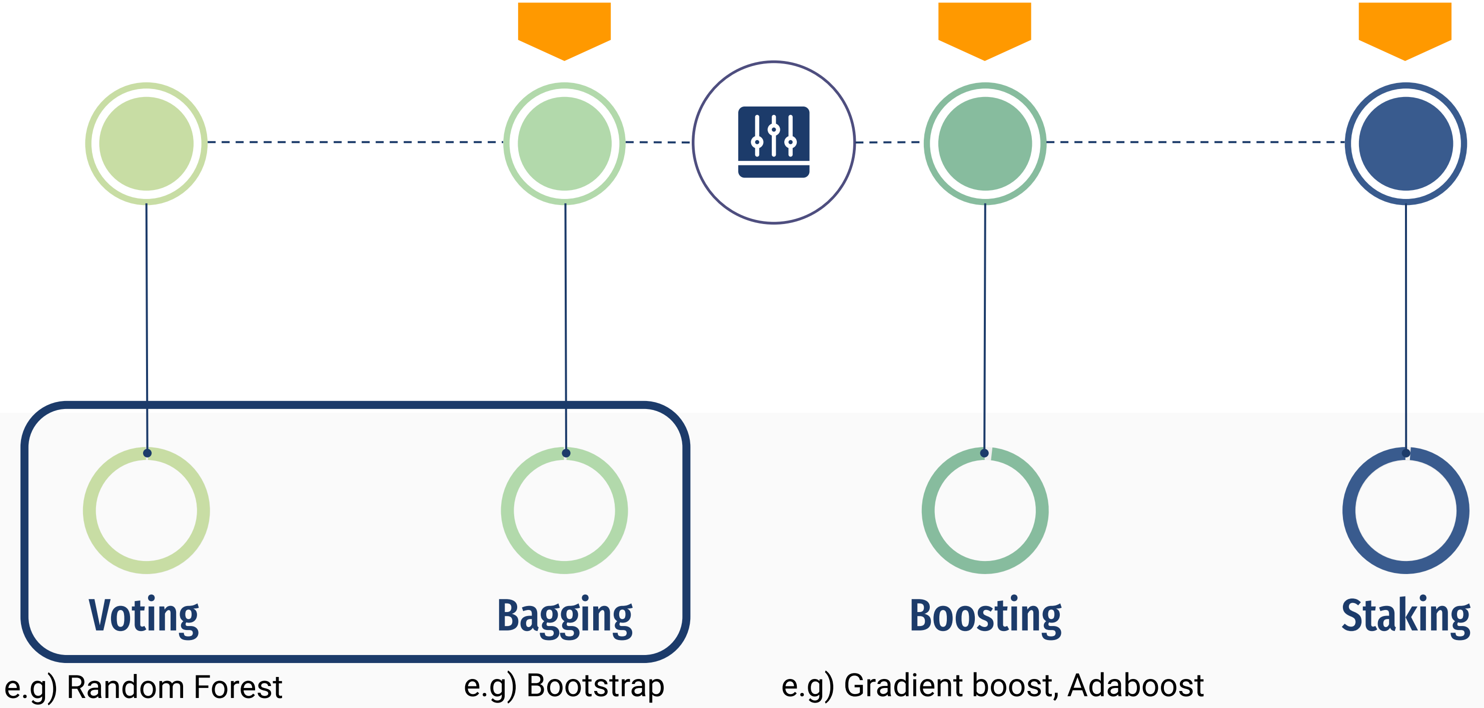
Ensemble Model

동일한 학습 알고리즘을 사용해 여러 모델을 학습하는 기법

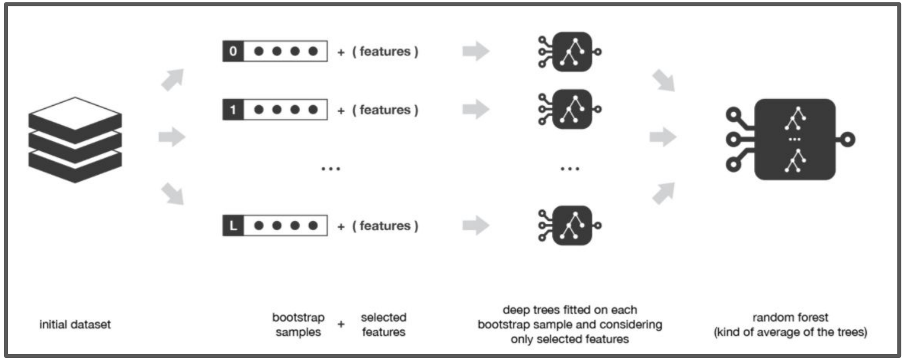
Single Learner(단일 학습기)보다 Weak Learner를 결합하면 더 좋은 성능을 얻을 수 있음

[Weak Learner 란? : 특정한 데이터를 이용해 인스턴스화 한 모델을 뜻하며, Weak Learner는 최종적인 결과물보다 상대적으로 정확하지 않은 결과를 보이는 learner]

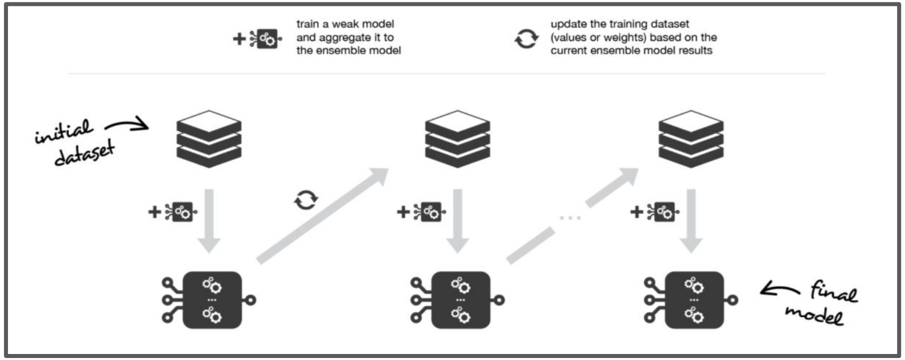
Ensemble Model Types



Bagging



Boosting



1

병렬적, 빠름

1. 기본 데이터를 샘플링하여, n개의 데이터셋을 만들어 n개의 모델을 학습시키고 **최종 결과를 aggregation**

2. 샘플링 후에는 n개의 모델이 독립적으로 동시에 각각의 데이터셋을 학습

3. 높은 bias로 인한 underfitting, 높은 Variance로 인한 overfitting 문제를 해결하는데 도움

2

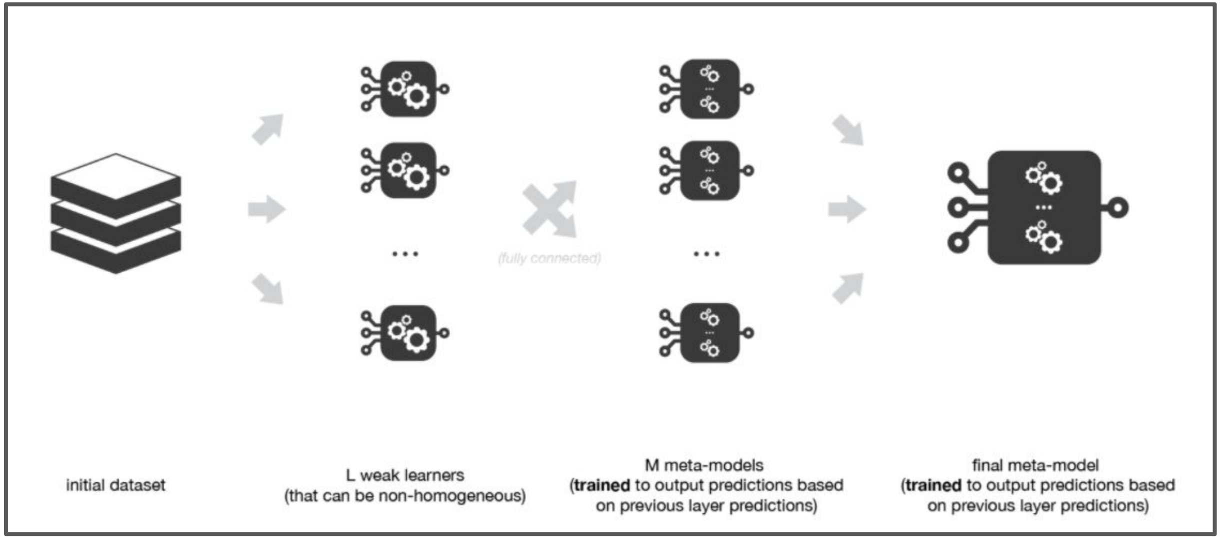
직렬적, 느림

1. 첫번째 모델이 기본 데이터셋을 그대로 학습하고, 다음 모델은 전체 데이터를 학습하되, 첫번째 모델이 맞추지 못한 데이터에 더 큰 중점을 두고 학습

2. Bagging에 비해 Boosting은 **맞추기 어려운 문제**를 맞추는데 특화, 앞 모델의 학습이 끝나야 뒷 모델이 **그 결과를 기반으로 가중치를 결정하고 학습**

3. Boosting의 경우 정확도가 높게 나타나지만 그만큼 Outlier에 취약

Staking



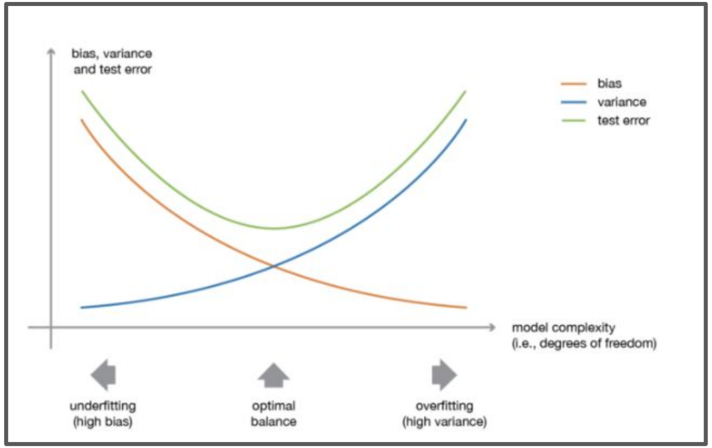
3

1. Bagging과 Boosting과 다르게 모델 별 예측된 값을 결합하여 최종 모델의 train data로 사용됨

2. Fold 별로 나누어진 데이터의 예측값과 최종모델의 예측값 비교하여 성능 평가

3. 주로 **모델의 성능을 올리기 위해서** 주로 사용됨**, 과적합 가능성 높음, 노력, 시간 다수 소요**

Summary



1. 단일 모형을 strong learner로 학습 시킬 경우 Bias, Variance의 차이를 해결하는데 어려움 존재

2. Weak learner는 Bias, Variance 줄이기 쉬움, 여러 weak learner를 결합하면 Bias, Variance가 낮은 Strong learner 생성 가능

3. 데이터에 따라 ensemble 모형을 사용할 경우 단일 모형보다 높은 성능을 기대할 수 있음

Application

01

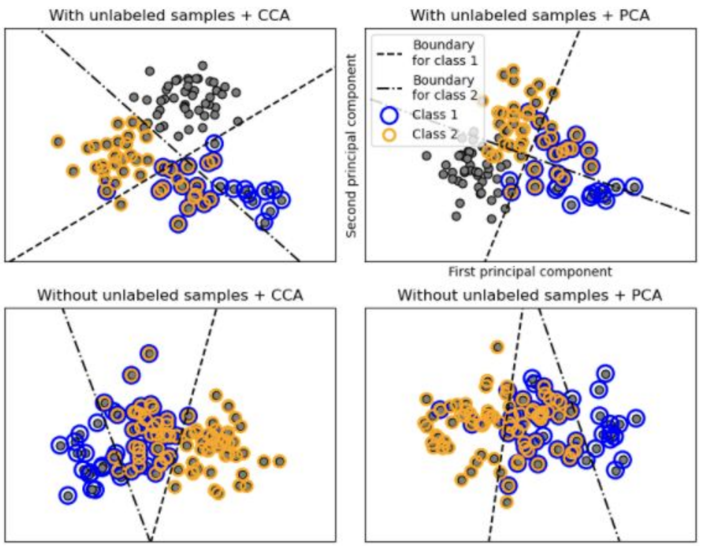
Multi classification 모델 사용하여 ensemble model로 구축할시 암종에 따른 다중 분류를 해결할 수 있고 정확도도 높을 것으로 사료됨

-> 4번째 논문 참조

02

Multi classification model은 deep learning에 속함
-> 많은 데이터 필요

Oversampling 사용하여 train data 증대하여 사용 가능, but 과적합 가능성 존재



Intensive Care Unit Mortality Prediction: An Improved Patient-Specific Stacking Ensemble Model

NORA EL-RASHIDY[1], SHAKER EL-SAPPAGH[2,3], TAMER ABUHMED[4], SAMIR ABDELRAZEK[5], AND HAZEM M. EL-BAKRY[5]

[1]Machine Learning and Information Retrieval Department, Faculty of Artificial Intelligence, Kafrelsheikh University, Kafr El-Sheikh 33516, Egypt
[2]Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain
[3]Information Systems Department, Faculty of Computers and Artificial Intelligence, Benha University, Banha 13518, Egypt
[4]College of Computing, Sungkyunkwan University, Seoul 561-758, South Korea
[5]Information Systems Department, Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt

TABLE 16. Score performance results for all the models.

Method	Algorithm	First 24 Hours				
		CV accuracy	F1	P	R	AUC
Score method	Apache II	-	0.721	0.8985	0.602	0.734
	SAPS- II	-	0.772	0.7720	0.767	0.812
	SOFA	-	0.733	0.752	0.708	0.782
Single Model	KNN	0.917±0.021	0.904	0.916	0.891	0.892
	DT	0.811±0.012	0.876	0.866	0.868	0.832
	LR	0.880±0.038	0.890	0.905	0.861	0.898
	LDA	0.901±0.098	0.902	0.894	0.872	0.891
	MLP	0.929±0.072	0.912	0.922	0.893	0.901
Ensemble Model	RF	0.867±0.015	0.776	0.899	0.752	0.842
	Bagging	0.922±0.007	0.901	0.940	0.903	0.916
	AdaBoost	0.935±1.105	0.911	0.933	0.900	0.904
	Voting	0.899±0.009	0.923	0.959	0.919	0.906
Proposed Model	Stacking	0.957±0.089	0.937	0.964	0.911	0.933

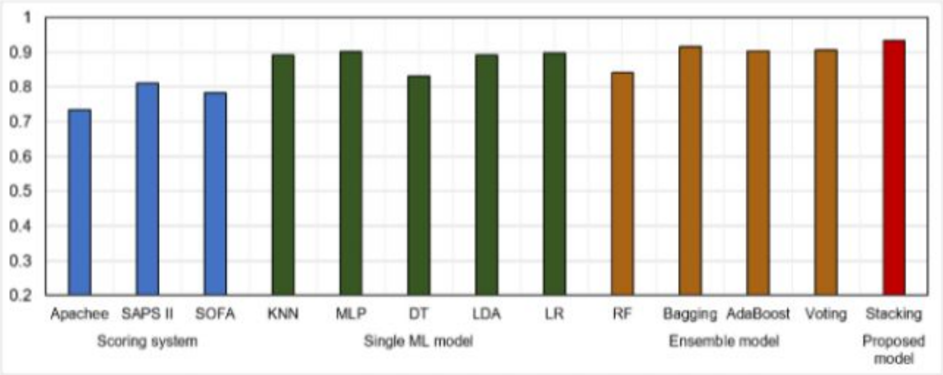


FIGURE 13. AUC scores of all the models evaluated in the study for a period of 24 hours.

Thank you