Artificial Intelligence in Medicine

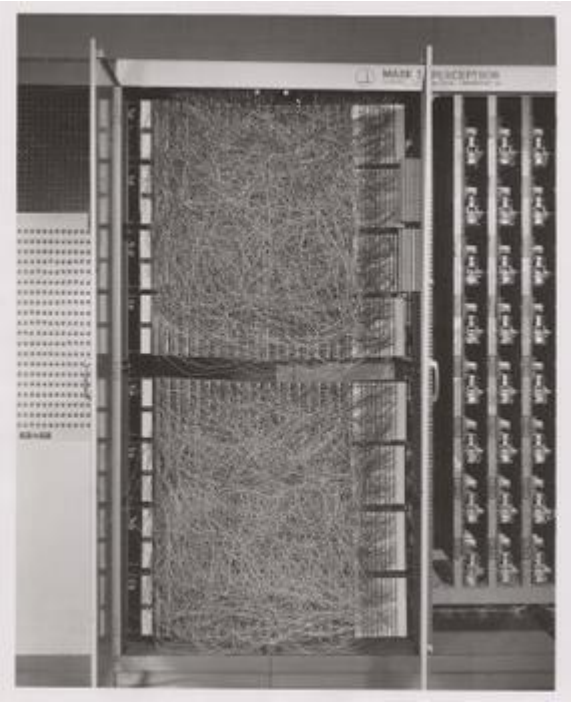
2019.4.27

vital@snu.ac.kr

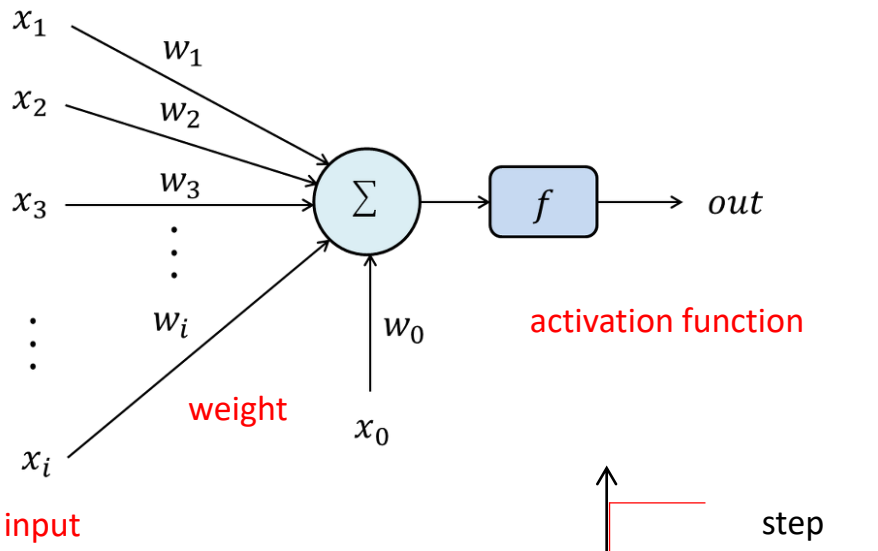
Seoul National University College of Medicine

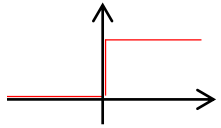
Department of Anesthesiology and Pain Medicine

Hyung-Chul Lee, MD, PhD

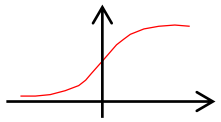


Perceptron, 1957





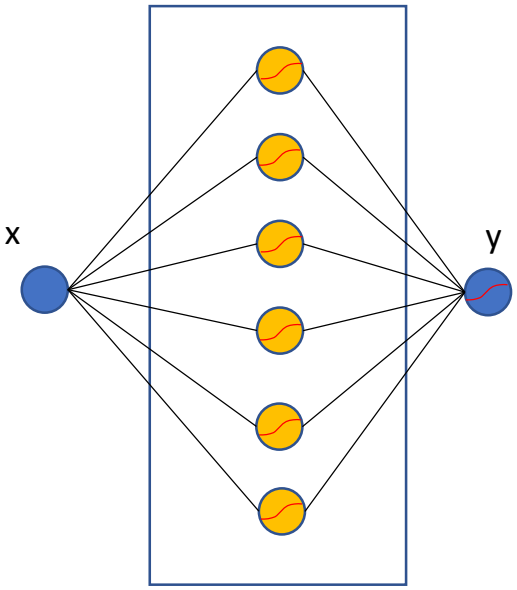
step



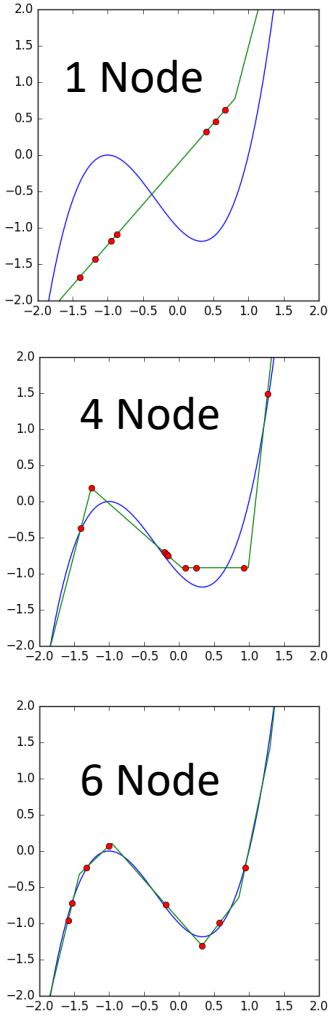
sigmoid

Multi-Layer Perceptron (MLP), 1980s

Universal approximation theorem, 1989

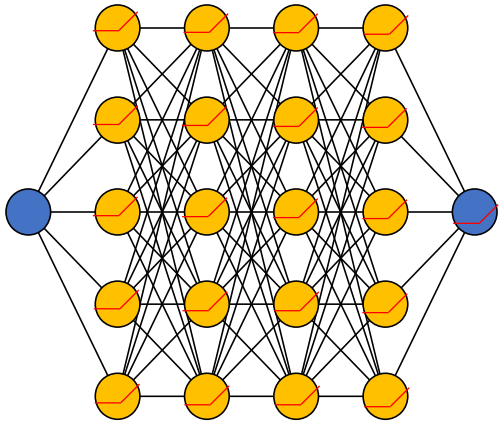


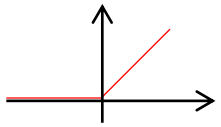
$$y = x^3 + x^2 - x - 1$$



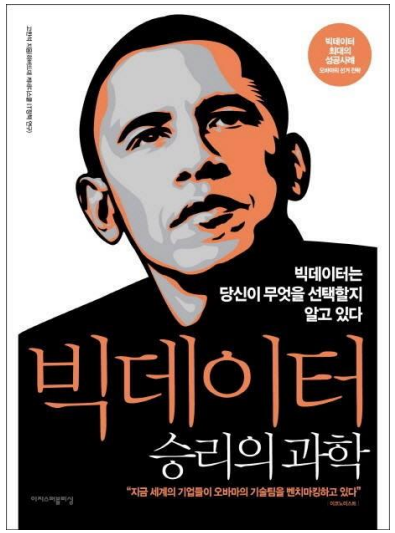
Deep Neural Network, 2000s

Big Data, 2010s





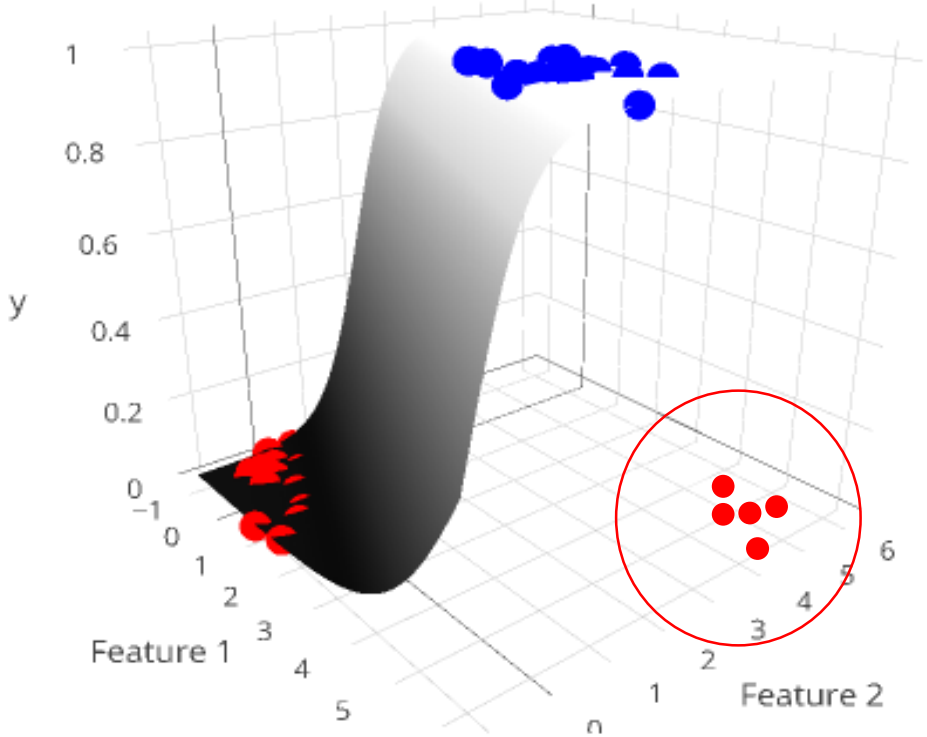
relu

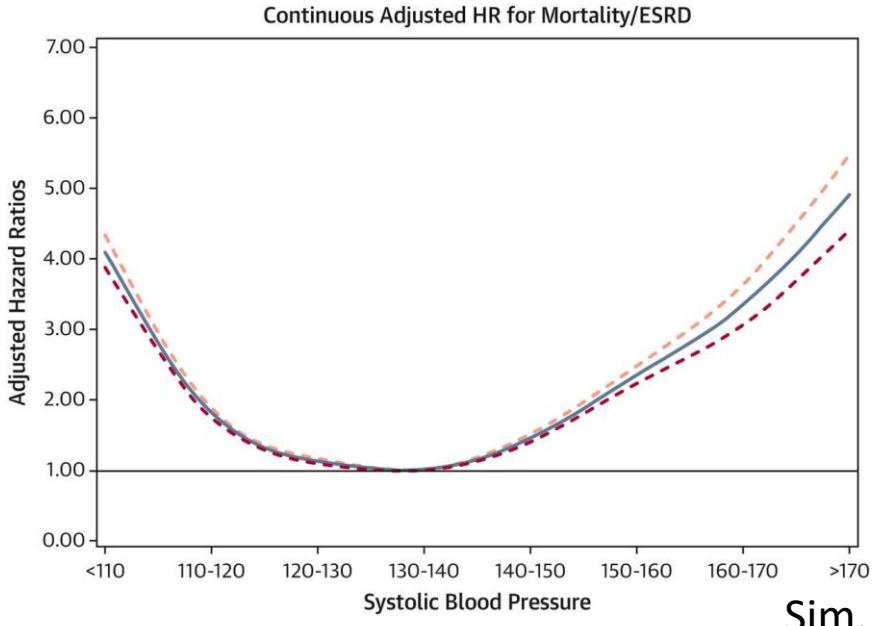


Geoffrey Hinton

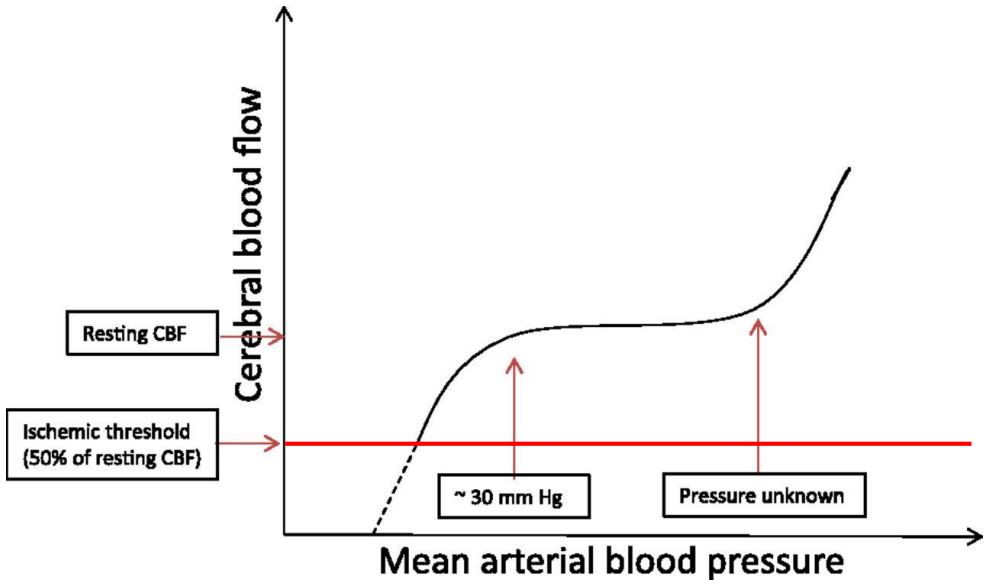
Logistic Regression

$$y = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 \ldots)}}$$





Sim, JACC, 2014



PERIOPERATIVE MEDICINE

Development and Validation of a Deep Neural Network Model for Prediction of Postoperative In-hospital Mortality

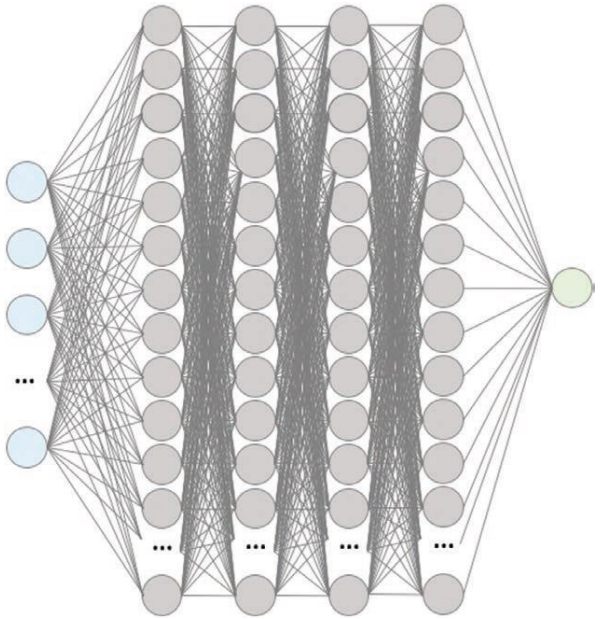
ANESTHESIOLOGY 2018; 129:649-62

Christine K. Lee, M.S., Ph.D., Ira Hofer, M.D., Eilon Gabel, M.D., Pierre Baldi, Ph.D., Maxime Cannesson, M.D., Ph.D.

Table 1. Eighty-seven Features Used in Models with Description and Applied Maximum Possible

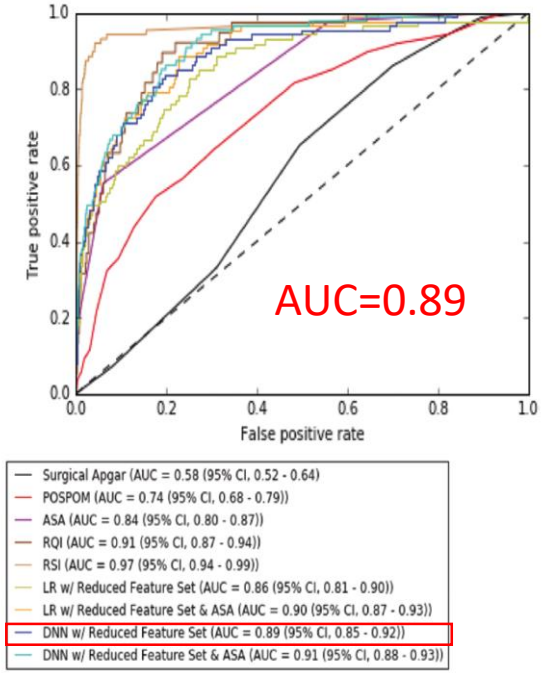
Feature Name(s)	Description	No Feat
COLLOID_ML*	Total colloid transfused (ml)	
CRYSTALLOID_ML*	Total crystalloid transfused (ml)	
DBP MAX*, MIN*, AVG, MED, STD	Maximum, minimum, average, median, and SD diastolic blood pressure for the case (mmHg)	
DBP_10min MAX, MIN, AVG, MED, STD	Maximum, minimum, average, median, and SD diastolic blood pressure for the last 10 min of the case (mmHg)	
EBL*	Total estimated blood loss (ml)	
EPHEDRINE BOLUS*	Total bolus dose of ephedrine (mg) during the case	
EPINEPHRINE BOLUS*, END RATE*, MAX RATE*	Total bolus dose (mcg), end of case infusion rate (mcg · kg^{-1} · min^{-1}), and highest infusion rate (mcg · kg^{-1} · min^{-1}) of epinephrine during the case	
ESMOLOL BOLUS*, END RATE*, MAX RATE*	Total bolus dose (mg), end of case infusion rate (mcg · kg^{-1} · min^{-1}), and highest infusion rate (mcg · kg^{-1} · min^{-1}) of esmolol during the case	
HR MAX*, MIN*, AVG, MED, STD	Maximum, minimum, average, median, and SD heart rate (beats/min) for the case	
HR_10min MAX, MIN, AVG, MED, STD	Maximum, minimum, average, median, and SD heart rate (beats/min) for the last 10 min of the case	
INVASIVE_LINE_YN*	Invasive central venous, arterial, or pulmonary arterial line used for the case (Yes/No)	
MAP MAX*, MIN*, AVG, MED, STD	Maximum, minimum, average, median, and SD mean blood pressure (mmHg) for the case	
MAP_10min MAX, MIN, AVG, MED, STD	Maximum, minimum, average, median, and SD mean blood pressure (mmHg) for the last 10 min of the case	
DES MAX*	Maximum and minimum alveolar concentration of desflurane during the case (note: this is not age adjusted)	
GLUCOSE MAX*, MIN*	Maximum and minimum plasma glucose concentration for the case (mg/dl)	
ISO MAX*	Maximum and minimum alveolar concentration of isoflurane during the case (note: this is not age adjusted)	
SEVO MAX*	Maximum and minimum alveolar concentration of sevoflurane during the case (note: this is not age adjusted)	
MILRINONE END RATE*, MAX RATE*	End of case infusion rate and highest infusion rate of milrinone during the case (mcg · kg^{-1} · min^{-1})	
HGB MIN*	Minimum hemoglobin concentration (g/dl) during the case	
MINUTES MAP < 50	Cumulative min with mean arterial pressure < 50 mmHg (min)	
MINUTES MAP < 60	Cumulative min with mean arterial pressure < 60 mmHg (min)	
NICARDIPINE END RATE*, MAX RATE*	End of case infusion rate and highest infusion rate of nicardipine during the case (mg/h)	
NITRIC_OXIDE_YN*	Nitric oxide used for the case (Yes/No)	
NITROGLYCERIN BOLUS*, END RATE*, MAX RATE*	Total bolus dose (mcg), end of case infusion rate (mcg/min), and highest infusion rate (mcg/min) of nitroglycerin during the case	
NITROPRUSSIDE END RATE*, MAX RATE*	End of case infusion rate and highest infusion rate of nitroprusside (mcg · kg^{-1} · min^{-1}) during the case	
PHENYLEPHRINE BOLUS*, END RATE*, MAX RATE*	Total bolus dose (mcg), end of case infusion rate (mcg/min), and highest infusion rate (mcg/min) of phenylephrine during the case	
SBP MAX*, MIN*, AVG, MED, STD	Maximum, minimum, average, median, and SD systolic blood pressure (mmHg) for the case	
SBP_10min MAX, MIN, AVG, MED, STD	Maximum, minimum, average, median, and SD systolic blood pressure (mmHg) for the last 10 min of the case	
Spo_2 MAX*, MIN*, AVG, MED, STD	Maximum, minimum, average, median, and SD Spo_2 (%) for the case	
Spo_2_10min MAX, MIN, AVG, MED, STD	Maximum, minimum, average, median, and SD Spo_2 (%) for the last 10 min of the case	
UOP*	Total urine output (ml)	
VASOPRESSIN BOLUS*, END RATE*, MAX RATE*	Total bolus dose (units), end of case infusion rate (units/h), and highest infusion rate (units/h) of vasopressin during the case	
XFUSION_RBC_ML*	Total red blood cells transfused (ml)	
	Total number of features	

*45 features used in the reduced feature set.



4 layers x 300 nodes

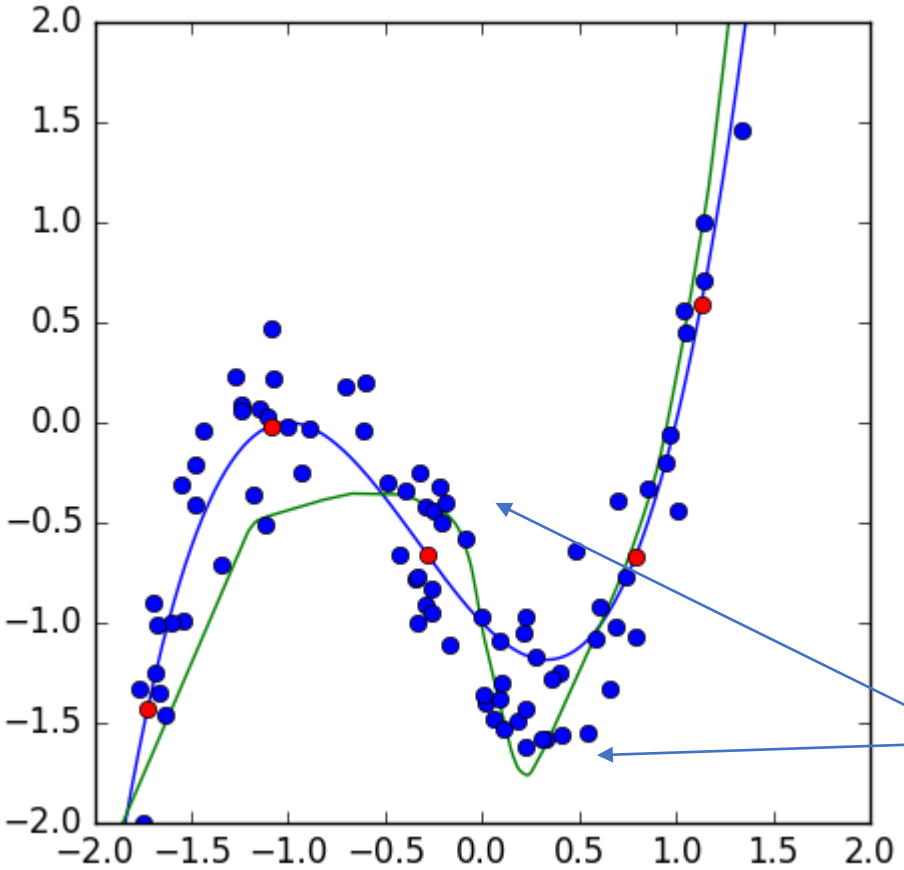
Input Vars	48 intraoperative variables
Output Var	In-hospital Mortality
Reference	Risk Stratification Index, Surgical Apgar, ASA class, ...
Method	FNN
Training Data	59,985 (Single center)
Test Data	20% (Randomly selected)
Performance	AUC=0.89 (cf. RSI 0.97)



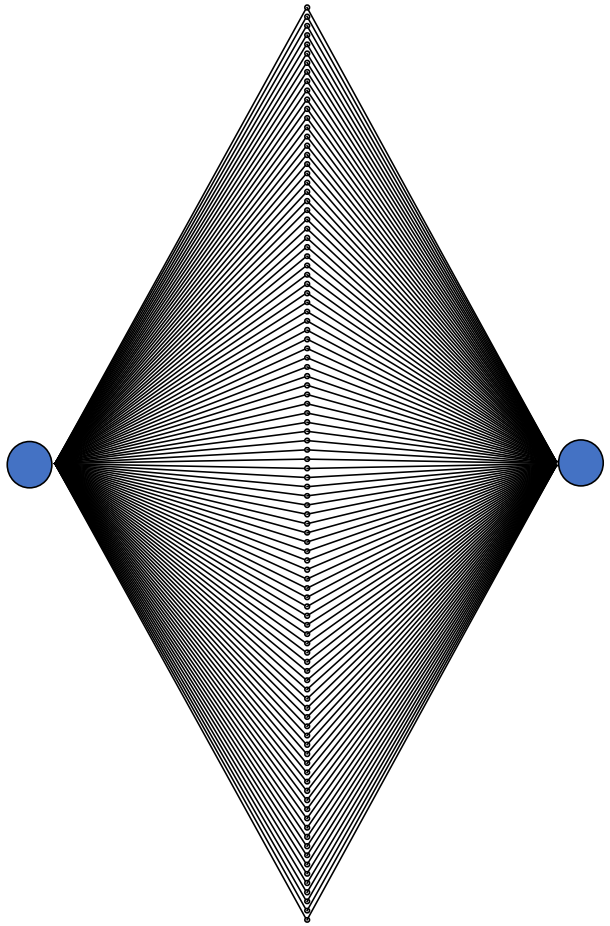
Overfitting

training samples

test samples

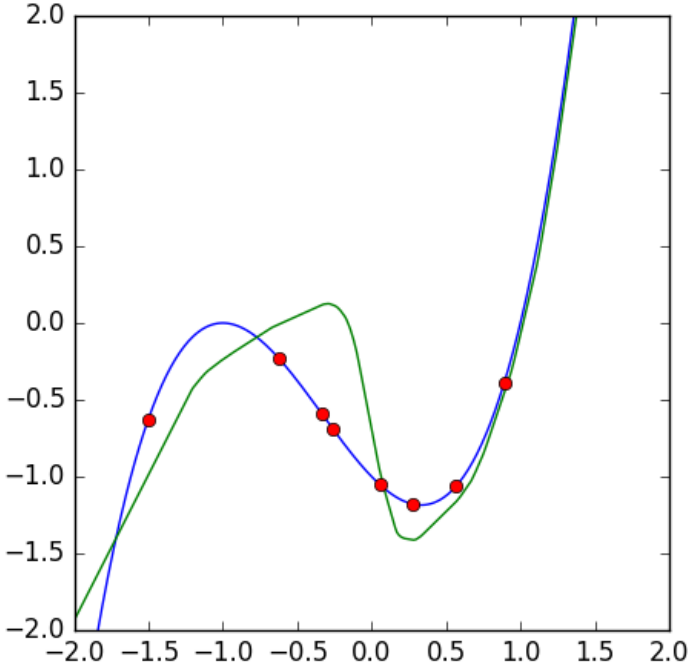


They learned noise!

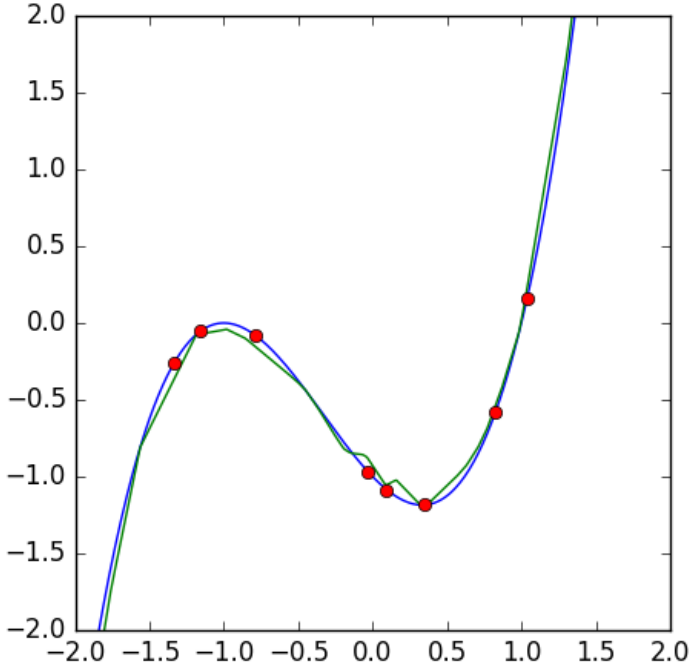


100 hidden nodes
100 training samples

More Data

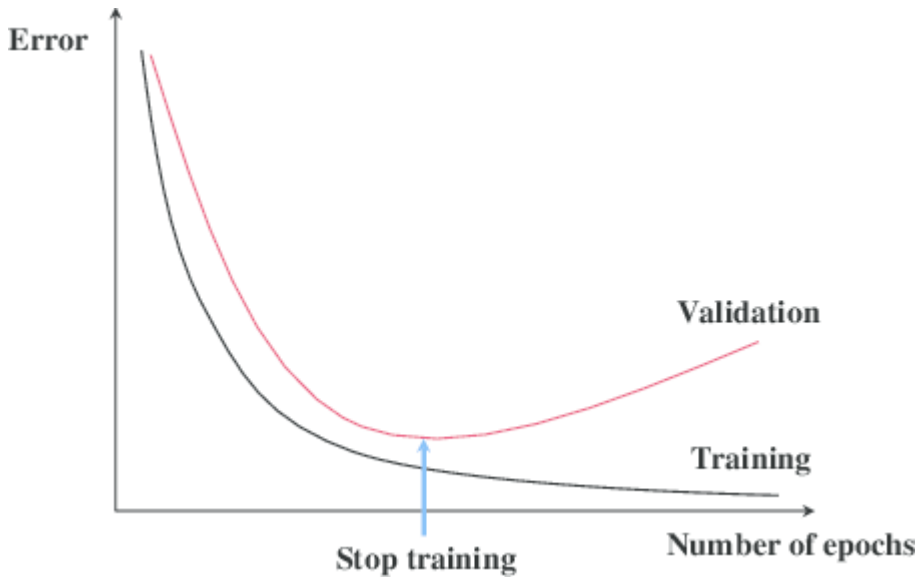


100 node
100 training samples
with noise size 0.1



100 node
1,000 training samples
with noise size 0.1

Early stopping

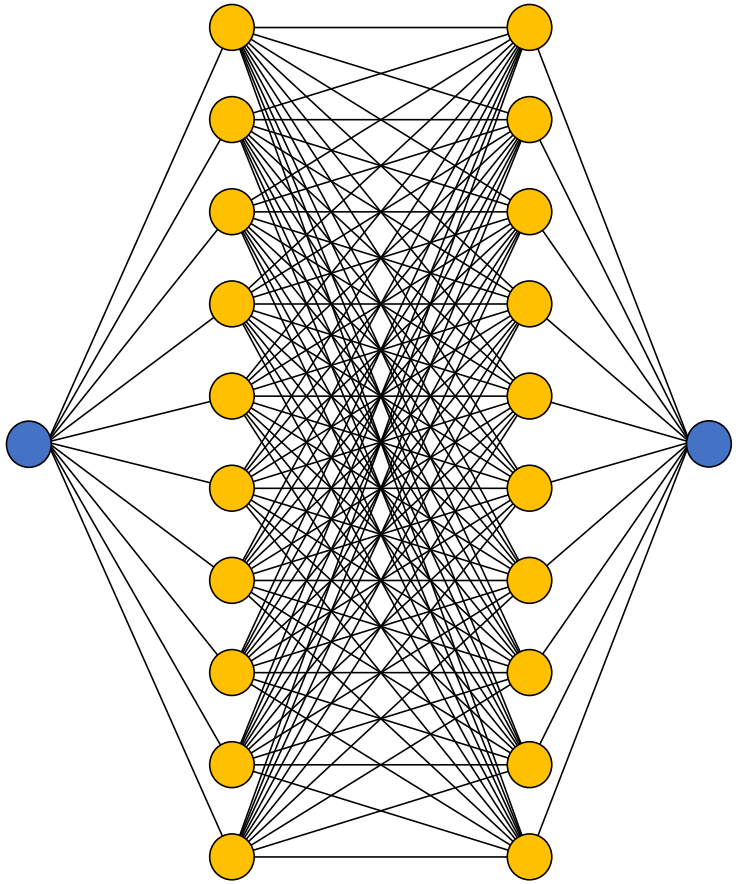


Training (60-70%)	Validation (10-20%)

Test (20%)

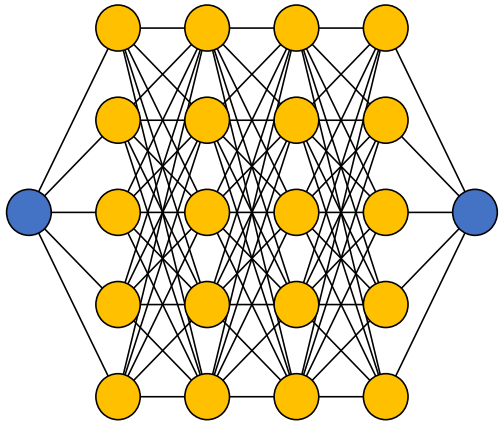
최종 모델 성능 평가용

Deeper Network



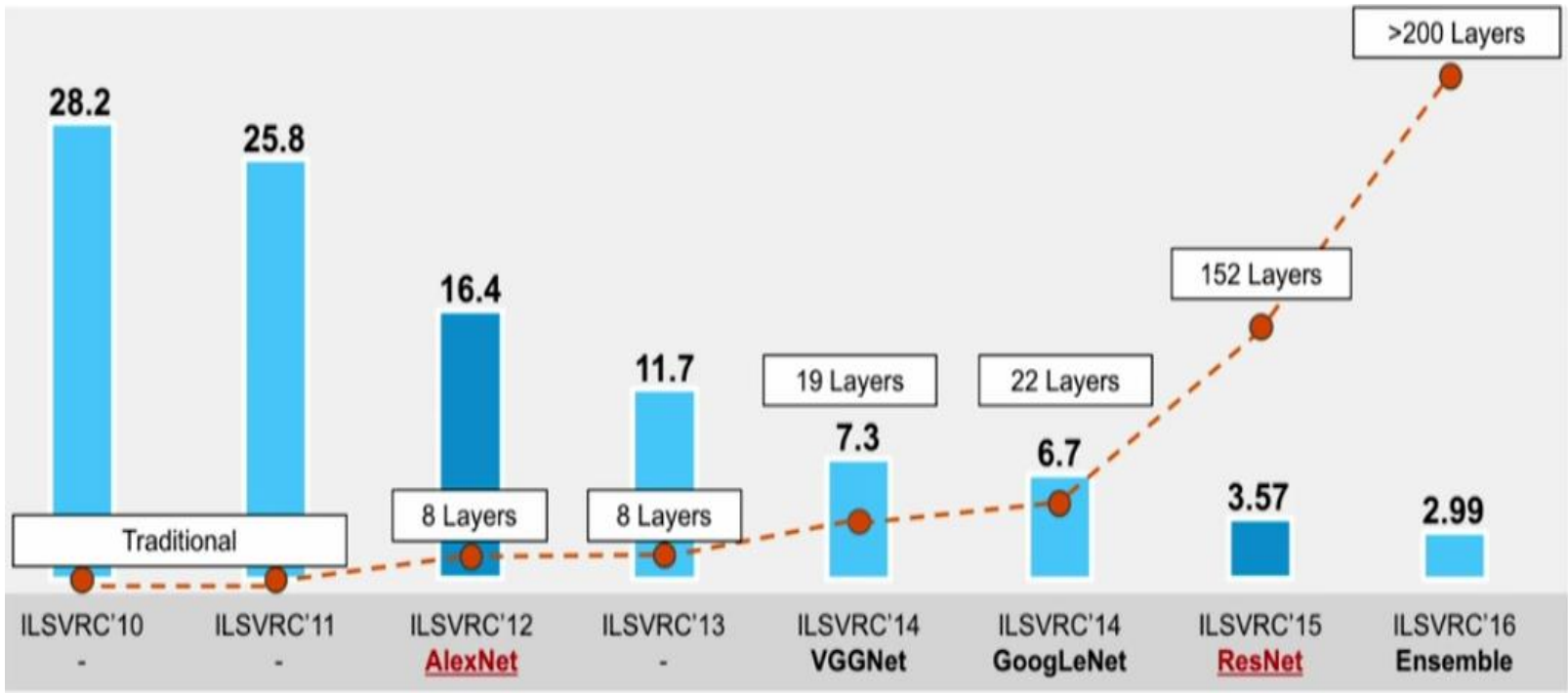
10 nodes x 2 layers

$10 + 10^2 + 10 = 120$ weights



5 nodes x 4 layers

$5 + 5^2 \times 3 + 5 = 85$ weights

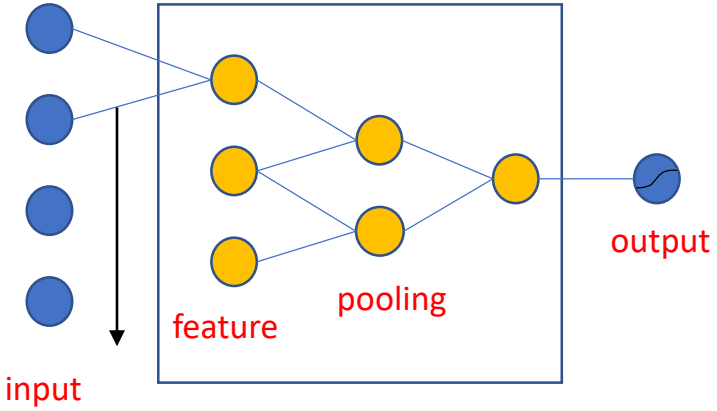


2014, Turing test

2016, AlphaGo

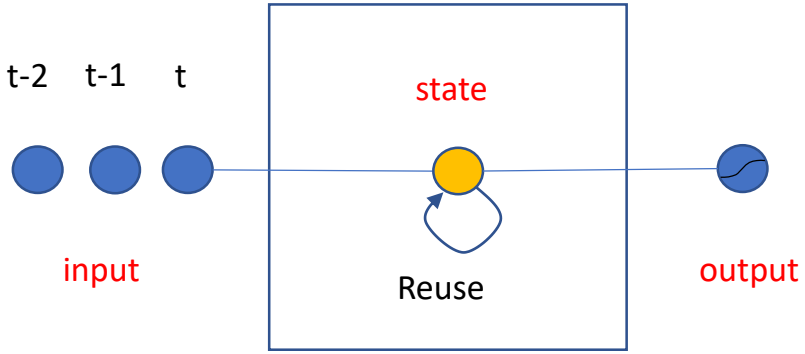
Better Networks

Concolutional Neural Network (CNN)

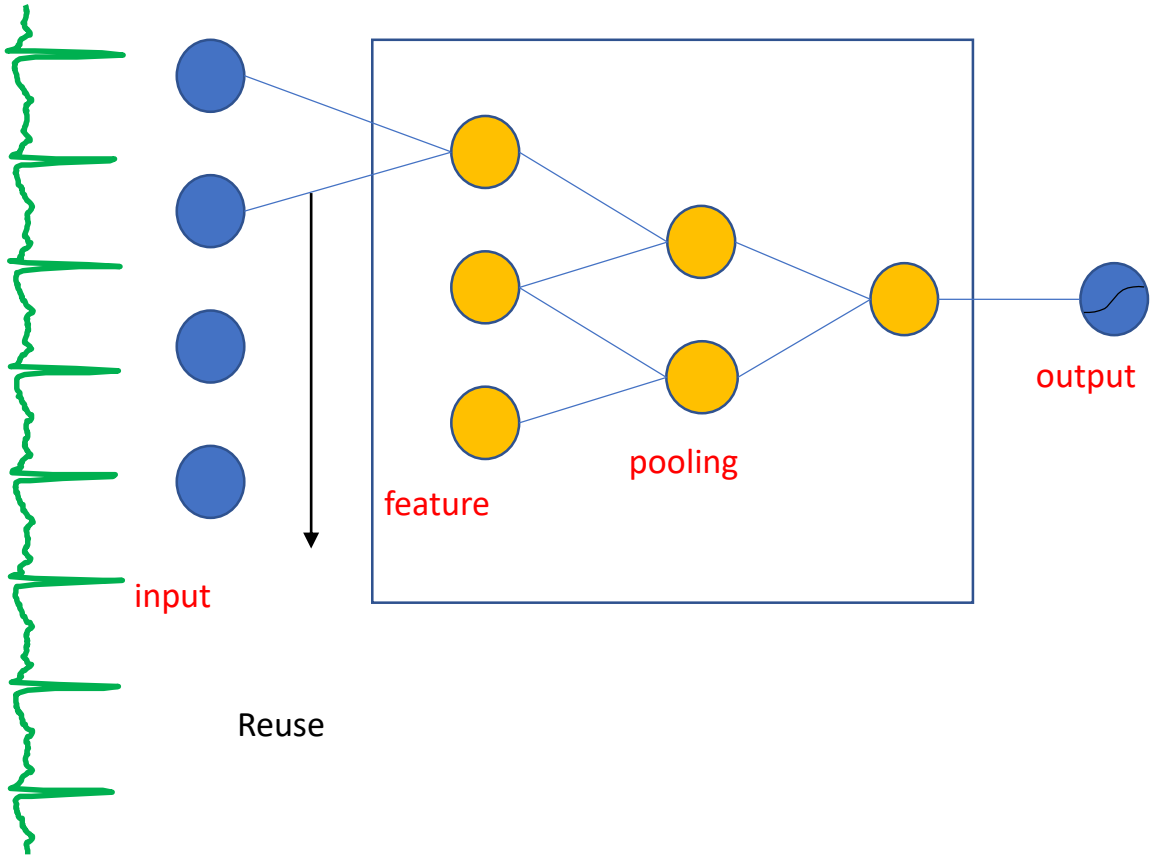


Reuse

Recurrent Neural Network (RNN)



Concolutional Neural Network (CNN)



2018.12, impact factor 33

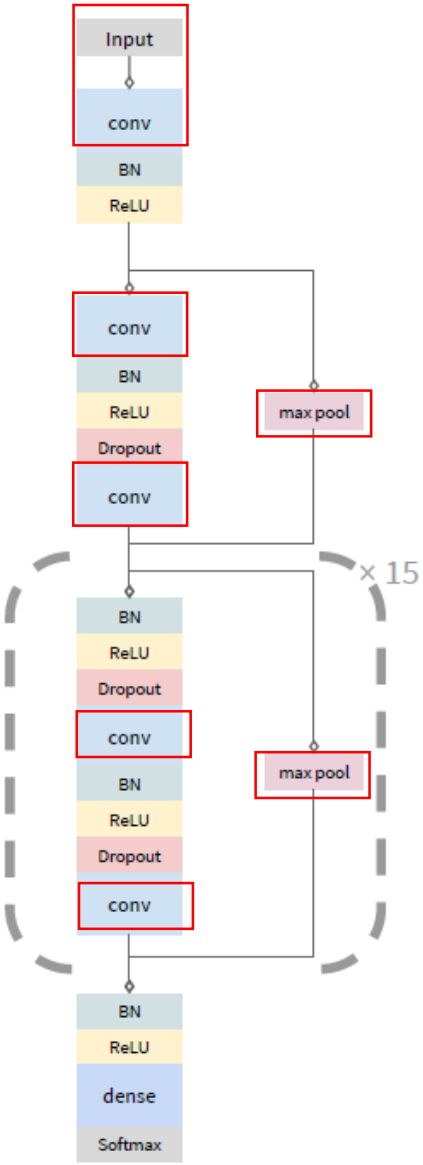
nature medicine FOCUS | LETTERS
https://doi.org/10.1038/s41591-018-0268-3
Corrected: Publisher Correction

Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network

Awni Y. Hannun[1,6]*, Pranav Rajpurkar[1,6], Masoumeh Haghpanahi[2,6], Geoffrey H. Tison[3,6], Codie Bourn[2], Mintu P. Turakhia[4,5] and Andrew Y. Ng[1]

	Algorithm AUC (95%
	Sequence[a]
Atrial fibrillation and flutter	0.973 (0.966–0.980)
AVB	0.988 (0.983–0.993)
Bigeminy	0.997 (0.991–1.000)
EAR	0.913 (0.889–0.937)
IVR	0.995 (0.989–1.000)
Junctional rhythm	0.987 (0.980–0.993)
Noise	0.981 (0.973–0.989)
Sinus rhythm	0.975 (0.971–0.979)
SVT	0.973 (0.960–0.985)
Trigeminy	0.998 (0.995–1.000)
Ventricular tachycardia	0.995 (0.980–1.000)
Wenckebach	0.978 (0.967–0.989)
Frequency-weighted average	0.978

ECG wave

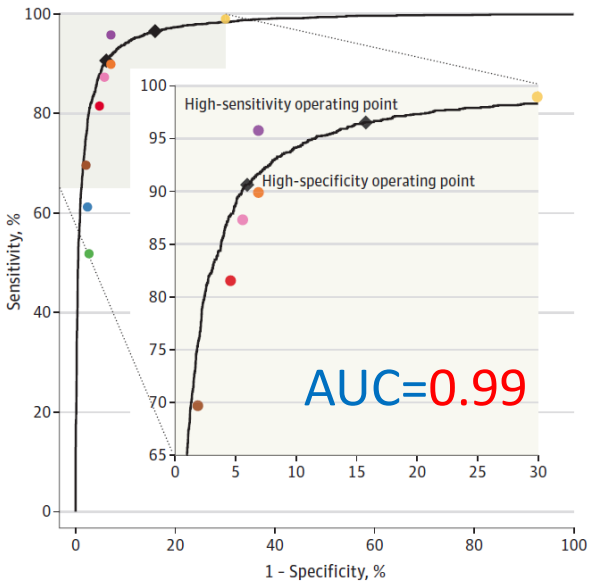


Afib, AVB, Bigeminy, Junctional, …

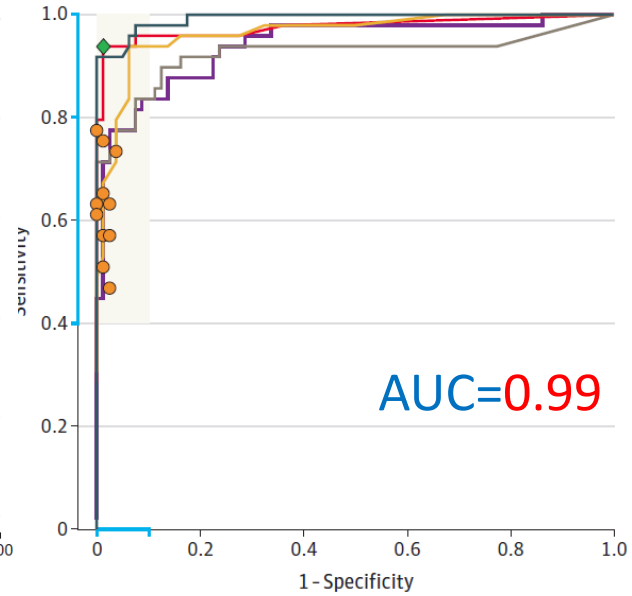
Input Vars	Single lead ECG waveform
Output Var	12 arrhythmia classification
Reference	Cardiologists
Method	CNN
Training Data	91,232 (Single center)
Test Data	328
Performance	AUC=0.97 F-score = 0.84 (cf. Cardiologists 0.78)

Andrew Ng

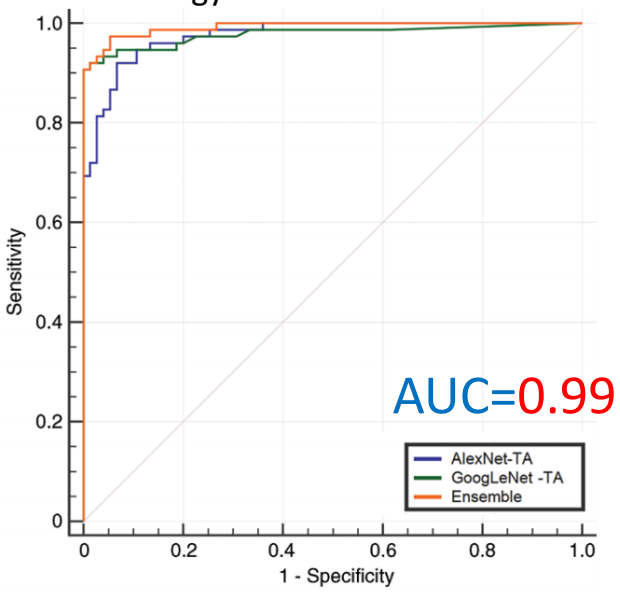
Ophthalmology



Pathology



Radiology



Gulshan, JAMA, 2017

Detection of Diabetic Retinopathy in Retinal Fundus Photographs

GoogLeNet+GoogLeNet

Bejnordi, JAMA, 2017

Detection of Lymph Node Metastases in Women With Breast Cancer

GoogLeNet

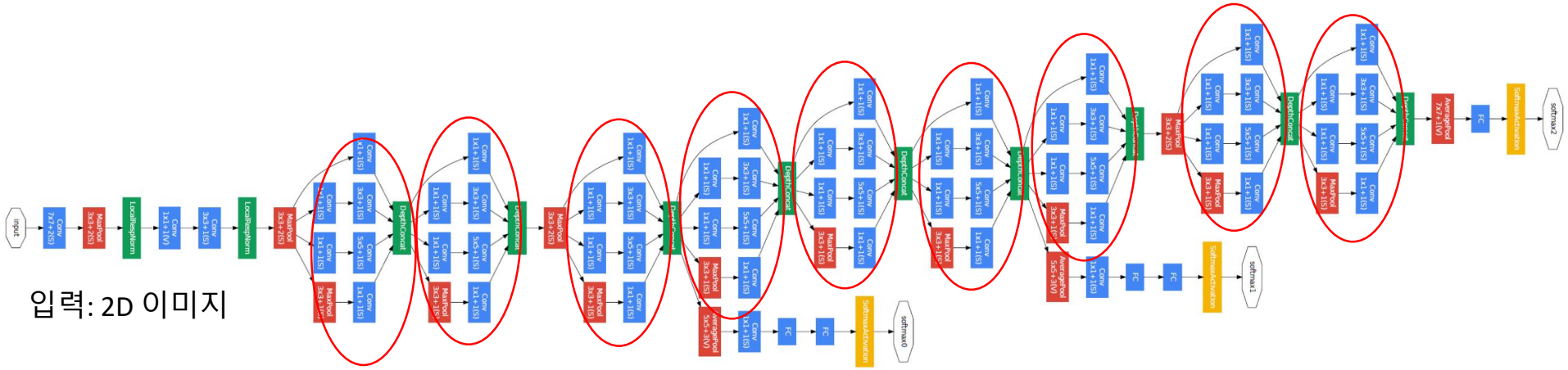
Lakhani, Radiology, 2017

Detection of Tuberculosis on chest radiographs

GoogLeNet:AlexNet=10:1

GoogLeNet

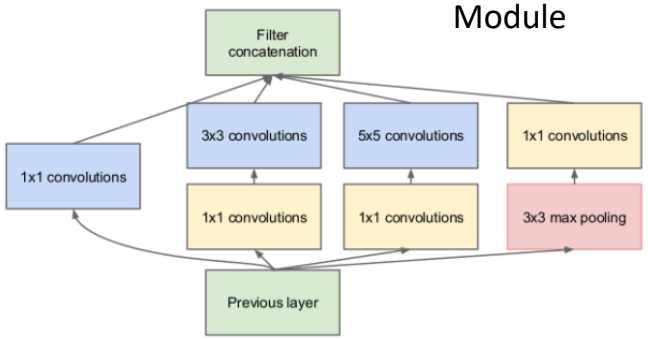
출력: 분류



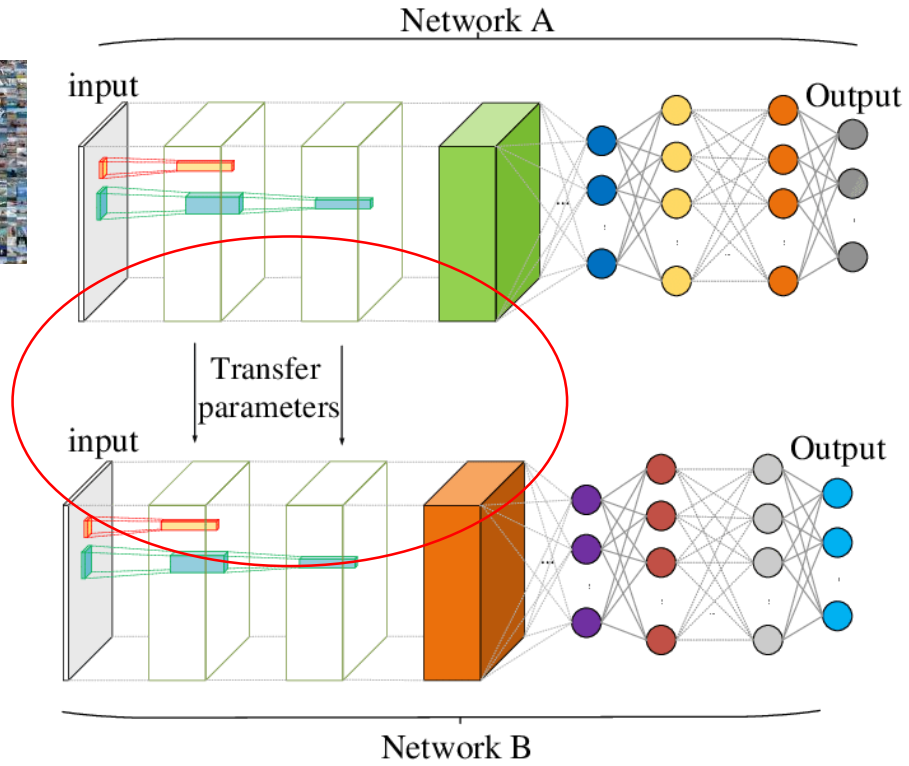
입력: 2D 이미지

Pretrained Weights

Inception Module



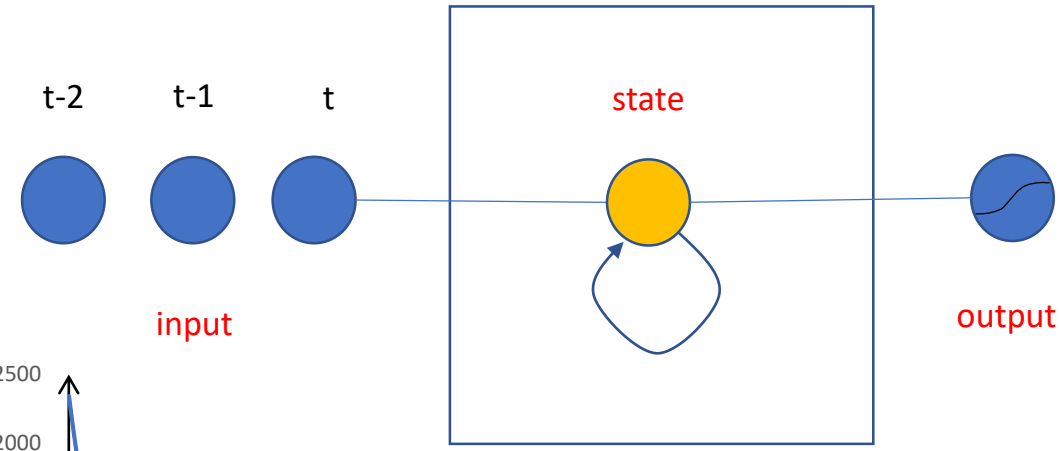
Transfer learning

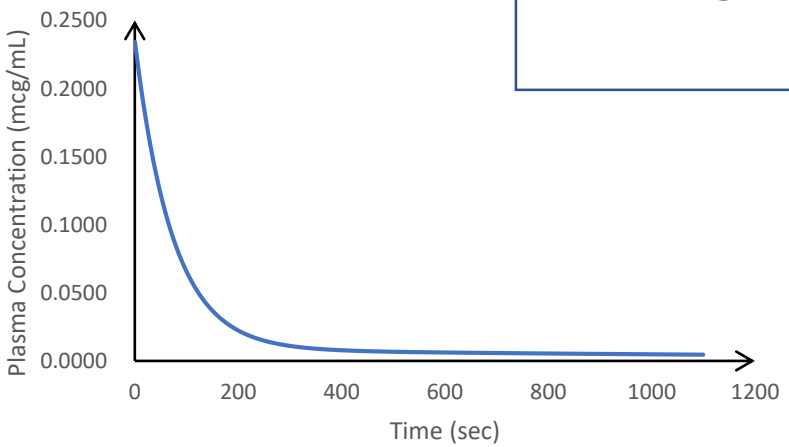


AUC Test Dataset

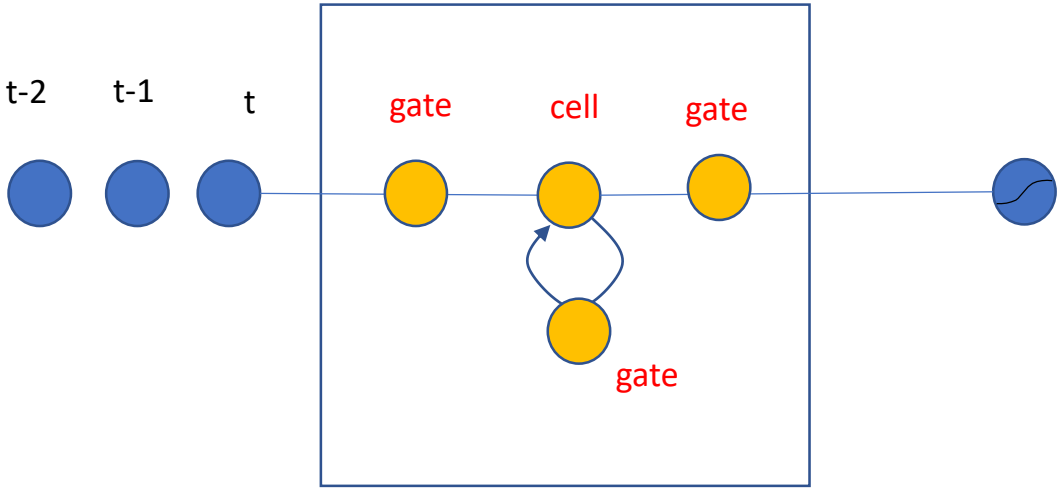
Parameter	Untrained	Pretrained	Untrained with Augmentation*	Pretrained with Augmentation*
AlexNet	0.90 (0.84, 0.95)	0.98 (0.95, 1.00)	0.95 (0.90, 0.98)	0.98 (0.94, 0.99)
GoogLeNet	0.88 (0.81, 0.92)	0.97 (0.93, 0.99)	0.94 (0.89, 0.97)	0.98 (0.94, 1.00)
Ensemble				0.99 (0.96, 1.00)

Recurrent Neural Network (RNN)





LSTM



THE LANCET
Respiratory Medicine

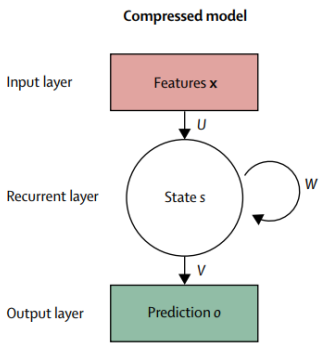
2018.9 Impact Factor **21**

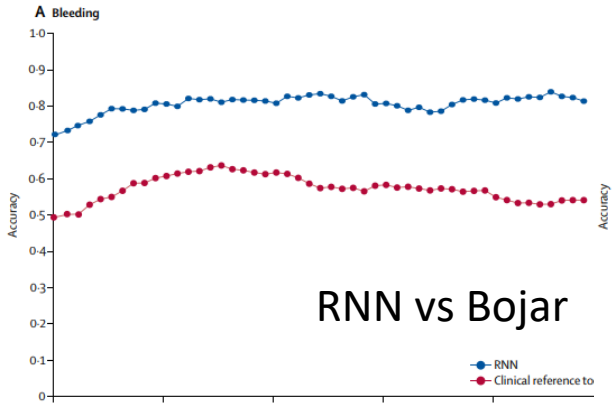
Machine learning for real-time prediction of complications in critical care: a retrospective study

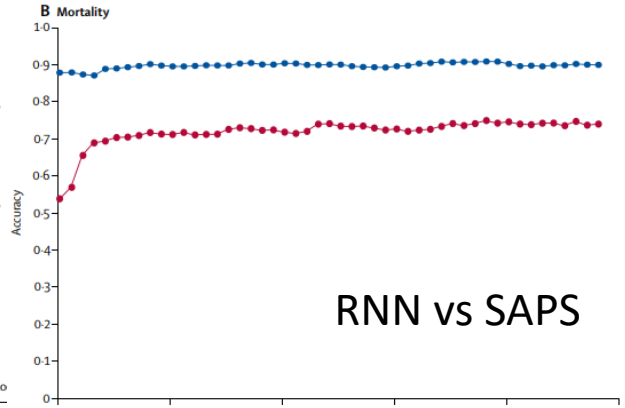
Alexander Meyer, Dina Zverinski, Boris Pfahringer, Jörg Kempfert, Titus Kuehne, Simon H Sündermann, Christof Stamm, Thomas Hofmann, Volkmar Falk, Carsten Eickhoff

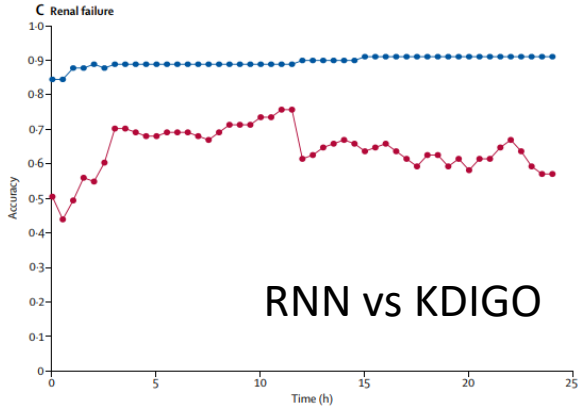
Patient information (four features)	Age, sex, height, weight
Information relating to initial surgery (nine features)	Anaesthesia type, American Society of Anesthesiologists Score, cardioplegic solution, aortic cross-clamp time, cardiopulmonary bypass time, anaesthetic monitoring time, surgery duration, surgery type, urgency
Vital signs (11 features)	Systolic, mean, and diastolic arterial pressure; systolic, mean, and diastolic pulmonary artery pressure; central venous pressure; ventilator FiO_2 setting; heart and respiratory frequency; body temperature
Arterial blood gas (nine features)	Bicarbonate, glucose, haemoglobin, oxygen saturation, partial pressure of carbon dioxide and oxygen, pH level, potassium, sodium
Laboratory results (17 features)	Albumin, bilirubin, urea, C-reactive protein, creatine kinase, γ-glutamyltransferase, glutamic oxaloacetic transaminase, haemoglobin , haematocrit, international normalised ratio, creatinine, white blood cell count, lactate dehydrogenase, magnesium, partial thromboplastin time, platelets, prothrombin time
Balance output (two features)	Bleeding rate, urine flow rate

Input Vars	52 features
Output Var	Bleeding, mortality, renal failure within 24-hours after surgery
Reference	Bleeding – Bojar Mortality – SAPS Renal Failure – KDIGO
Method	RNN
Training Data	11,492 (Single center)
Test Data	53,423 (MIMIC-III)
Performance	Bleeding AUC=0.87 (cf. 0.58) Mortality AUC=0.95 (cf. 0.71) Renal Failure AUC=0.96 (cf. 0.72)



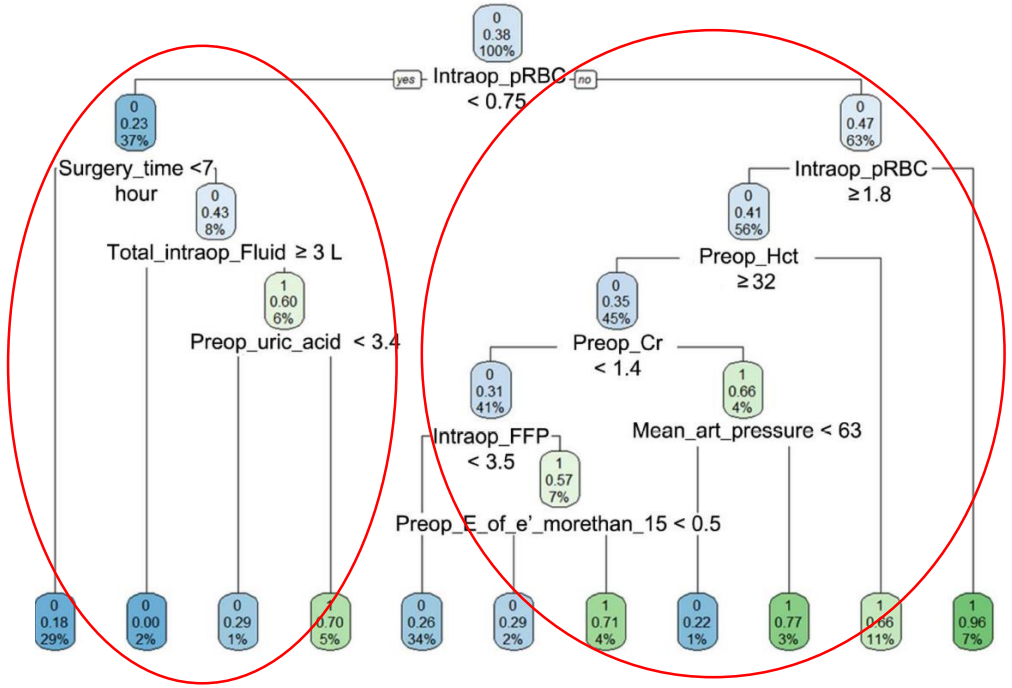


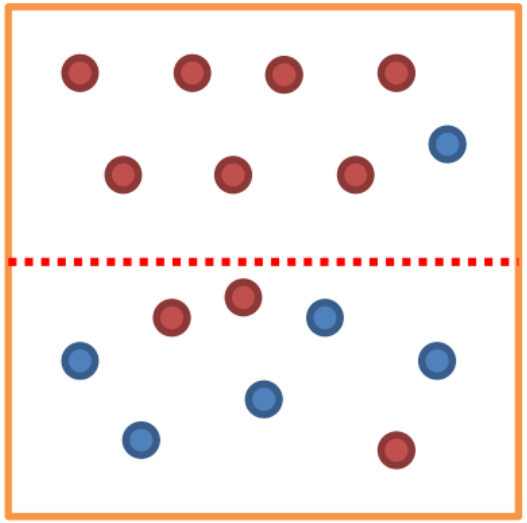




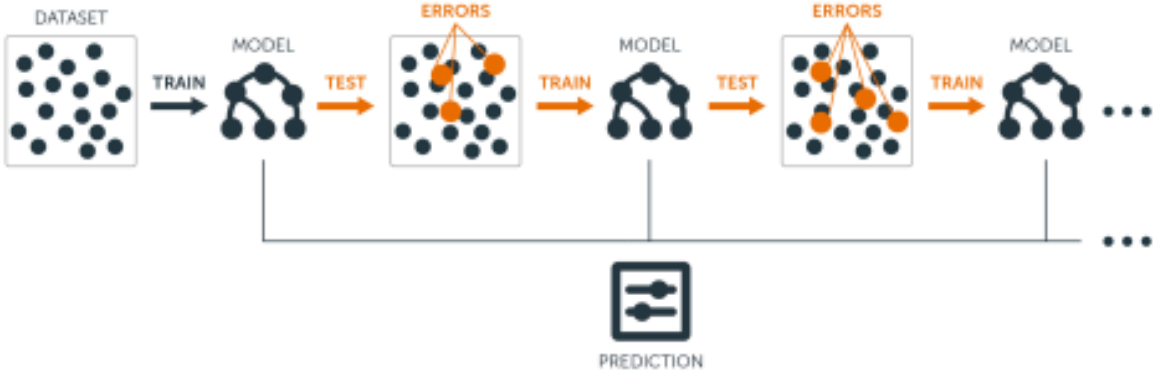
Gradient Boosting Machine

Decision Tree



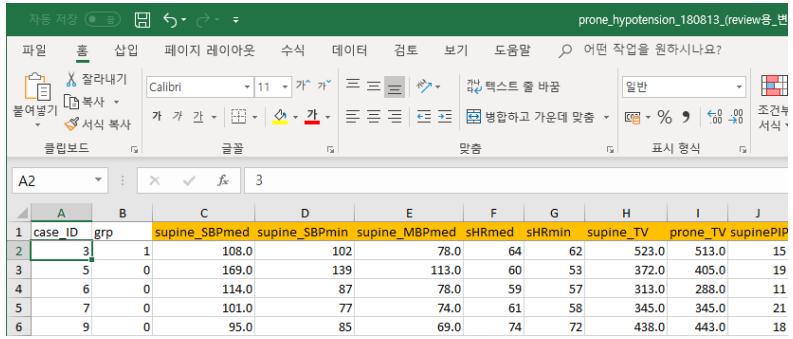


Gradient Boosting Machine

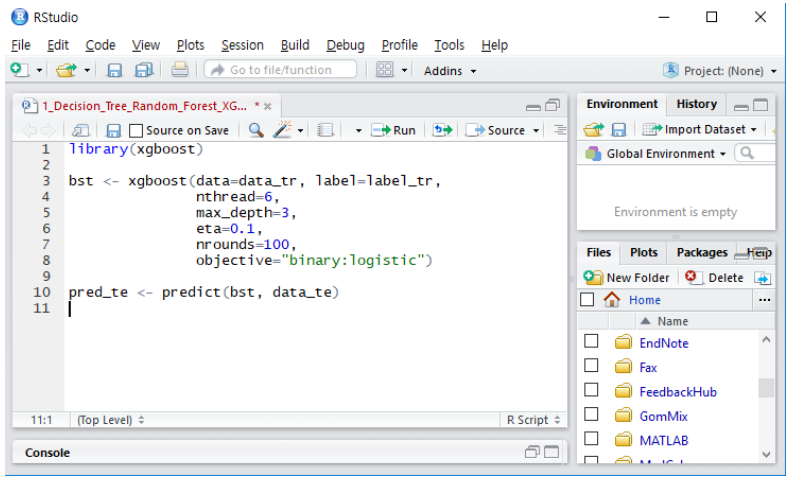


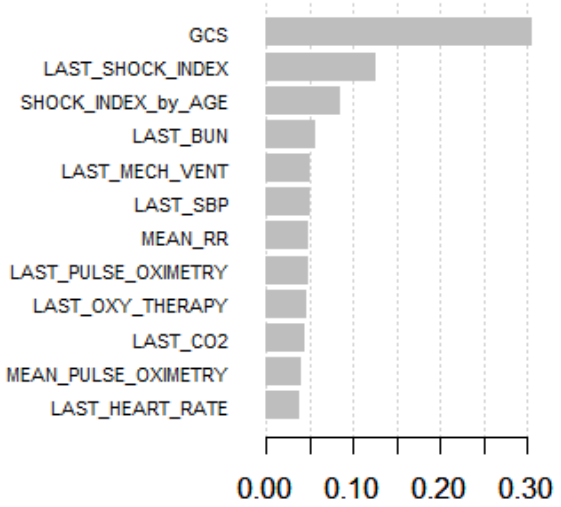
$$y = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 \ldots)}}$$

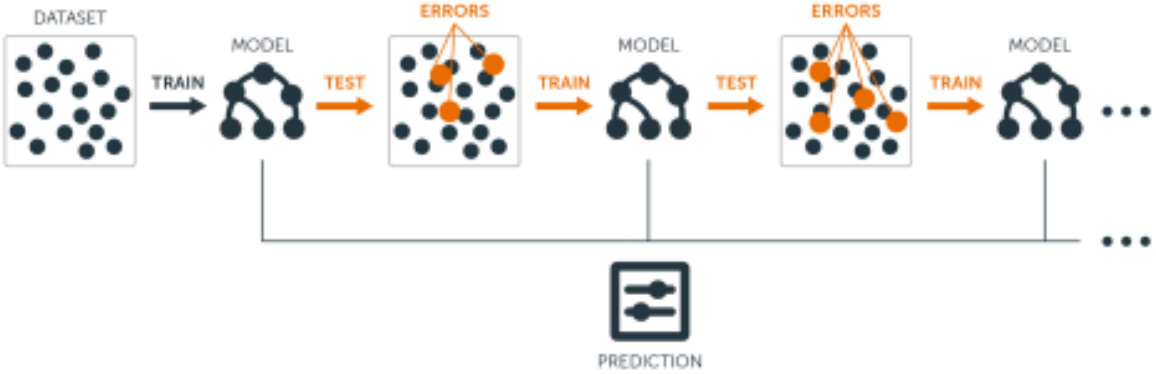
Gradient Boosting Machine

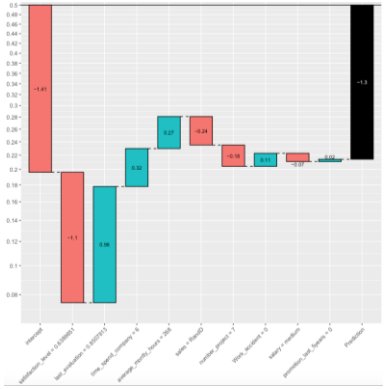


case_ID	grp	supine_SBPmed	supine_SBPmin	supine_MBPmed	sHRmed	sHRmin	supine_TV	prone_TV	supinePIP
3	1	108.0	102	78.0	64	62	523.0	513.0	15
5	0	169.0	139	113.0	60	53	372.0	405.0	19
6	0	114.0	87	78.0	59	57	313.0	288.0	11
7	0	101.0	77	74.0	61	58	345.0	345.0	21
9	0	95.0	85	69.0	74	72	438.0	443.0	18









CCM 2018

Development and Evaluation of an Automated Machine Learning Algorithm for In-Hospital Mortality Risk Adjustment Among Critical Care Patients*

Ryan J. Delahanty, PhD[1]; David Kaufman, MD, FCCM[2]; Spencer S. Jones, PhD[1]

Objectives: Risk adjustment algorithms for ICU mortality are necessary for measuring and improving ICU performance. Existing risk adjustment algorithms are not widely adopted. Key barriers to adoption include licensing and implementation costs as well as labor costs associated with human-intensive data collection. Widespread adoption of electronic health records makes automated risk adjustment feasible. Using modern machine learning methods and open source tools, we developed and evaluated a retrospective risk adjustment algorithm for in-hospital mortality among ICU patients. The Risk of Inpatient Death score can be fully automated and is reliant upon data elements that are generated in the course of usual hospital processes.

Setting: One hundred thirty-one ICUs in 53 hospitals operated by Tenet Healthcare.

Patients: A cohort of 237,173 ICU patients discharged between January 2014 and December 2016.

Design: The data were randomly split into training (36 hospitals), and validation (17 hospitals) data sets. Feature selection and model training were carried out using the training set while the discrimination, calibration, and accuracy of the model were assessed in the validation data set.

Measurements and Main Results: Model discrimination was evaluated based on the area under receiver operating characteristic scores and visual analysis of calibration curves. Seventeen features, including a mix of clinical and administrative data elements, were retained in the final model. The Risk of Inpatient Death score demonstrated excellent discrimination (area under receiver operating characteristic curve = 0.94) and calibration (adjusted Brier score = 52.8%) in the validation dataset; these results compare favorably to the published performance statistics for the most commonly used mortality risk adjustment algorithms.

Conclusions: Low adoption of ICU mortality risk adjustment algorithms impedes progress toward increasing the value of the healthcare delivered in ICUs. The Risk of Inpatient Death score has many attractive attributes that address the key barriers to adoption of ICU risk adjustment algorithms and performs comparably to existing human-intensive algorithms. Automated risk adjustment algorithms have the potential to obviate known barriers to adoption such as cost-prohibitive licensing fees and significant direct labor costs. Further evaluation is needed to ensure that the level of performance observed in this study could be achieved at independent sites. (*Crit Care Med* 2018; 46:e481–e488)

Key Words: critical care; ICU scoring systems; machine learning; mortality risk

ICU mortality prediction

AUC 0.94 *cf.* APACHE score (0.8)

Input Vars	17 features (lab, vital signs)
Output Var	In-hospital Mortality
Reference	(-)
Method	Gradient Boosting Machine
Training Data	146,982 (36 hospitals) 90,191 (17 hospitals)
Test Data	17 hospitals
Performance	AUC=0.94 (17 features) AUC=0.91 (14 features)

Feature	Mean (SD)	Missingness	Relative influence
Last Glasgow Coma Score (integer 1-15)	13.4 (3)	8.98%	30.54%
Last shock index	0.7 (0.3)	0.26%	12.51%
Last measured shock index (HR/SBP) x age	43.7 (21.1)	0.26%	8.59%
Last blood urea nitrogen	25.1 (20.4)	3.54%	5.62%
Last mechanical ventilation status (Y/N)	14.4% (Y)	3.96%	5.18%
Last systolic blood pressure	123.5 (23.9)	0.15%	5.08%
Mean respiratory rate	19.2 (3.8)	0.10%	4.88%
Last pulse oximetry	96.1 (7.4)	0.44%	4.80%
Last evidence of any oxygen therapy (Y/N)	59.2% (Y)	3.96%	4.74%
Last CO2 measurement	24.5 (4.7)	3.26%	4.54%
Mean pulse oximetry	97 (2.6)	0.44%	4.09%
Last heart rate	82.4 (20.1)	0.08%	3.84%
Mean temperature Fahrenheit	98.3 (0.8)	3.75%	3.22%
Change in creatinine level	-0.14 (0..83)	3.59%	2.37%

The Development of a Machine Learning Inpatient Acute Kidney Injury Prediction Model*

Jay L. Koyner, MD; Kyle A. Carey, MPH; Dana P. Edelson, MD, MS; Matthew M. Churpek, MD, MPH, PhD

JCM 2018

Objectives: To develop an acute kidney injury risk prediction model using electronic health record data for longitudinal use in hospitalized patients.
Design: Observational cohort study.
Setting: Tertiary, urban, academic medical center from November 2008 to January 2016.
Patients: All adult inpatients without pre-existing renal failure at admission, defined as first serum creatinine greater than or equal to 3.0 mg/dL, *International Classification of Diseases*, 9th Edition, code for chronic kidney disease stage 4 or higher or having received renal replacement therapy within 48 hours of first serum creatinine measurement.
Interventions: None.
Measurements and Main Results: Demographics, vital signs, diagnostics, and interventions were used in a Gradient Boosting Machine algorithm to predict serum creatinine–based Kidney Disease Improving Global Outcomes stage 2 acute kidney injury, with 60% of the data used for derivation and 40% for validation. Area under the receiver operator characteristic curve (AUC) was calculated in the validation cohort, and subgroup analyses were conducted across admission serum creatinine, acute kidney injury severity, and hospital location. Among the 121,158 included patients, 17,482 (14.4%) developed any Kidney Disease Improving Global Outcomes acute kidney injury, with 4,251 (3.5%) developing stage 2. The AUC (95% CI) was 0.90 (0.90–0.90) for predicting stage 2 acute kidney injury within 24 hours and 0.87 (0.87–0.87) within 48 hours. The AUC was 0.96 (0.96–0.96) for receipt of renal replacement therapy (n = 821) in the next 48 hours. Accuracy was similar across hospital settings (ICU, wards, and emergency department) and admitting serum creatinine groupings. At a probability threshold of greater than or equal to 0.022, the algorithm had a sensitivity of 84% and a specificity of 85% for stage 2 acute kidney injury and predicted the development of stage 2 a median of 41 hours (interquartile range, 12–141 hr) prior to the development of stage 2 acute kidney injury.
Conclusions: Readily available electronic health record data can be used to predict impending acute kidney injury prior to changes in serum creatinine with excellent accuracy across different patient locations and admission serum creatinine. Real-time use of this model would allow early interventions for those at high risk of acute kidney injury. (*Crit Care Med* 2018; 46:1070–1077)
Key Words: acute kidney injury; biomarker; electronic health record; risk assessment

Acute kidney injury (AKI) is a common clinical syndrome in hospitalized patients and is associated with increased costs, morbidity, and mortality (1–3). AKI is defined by either an increase in serum creatinine (SCr) or decreases in urine output according to consensus definitions established over the last decade (4). It is hypothesized that earlier detection of AKI may improve patient

AKI prediction

AUC 0.90

Article

Prediction of Acute Kidney Injury after Liver Transplantation: Machine Learning Approaches vs. Logistic Regression Model

CCM 2018

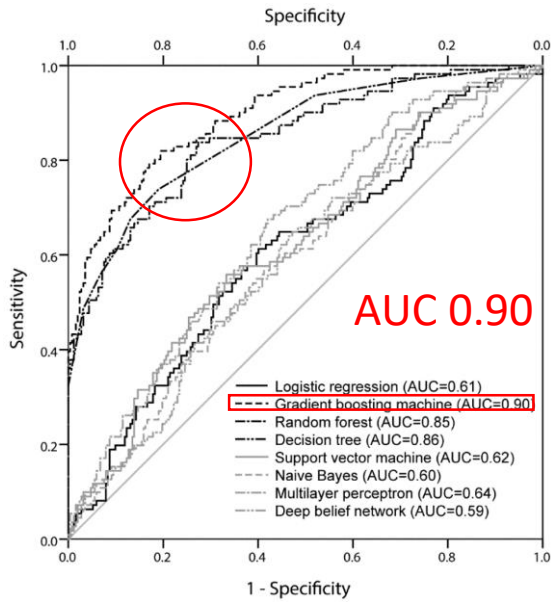
Hyung-Chul Lee [1], Soo Bin Yoon [1], Seong-Mi Yang [1], Won Ho Kim [1,2,*], Ho-Geol Ryu [1,2], Chul-Woo Jung [1,2], Kyung-Suk Suh [3] and Kook Hyun Lee [1,2]

Article

Derivation and Validation of Machine Learning Approaches to Predict Acute Kidney Injury after Cardiac Surgery

CCM 2018

Hyung-Chul Lee, Hyun-Kyu Yoon, Karam Nam, Youn Joung Cho, Tae Kyong Kim, Won Ho Kim * and Jae-Hyon Bahk



Surgery 2018

Postoperative bleeding risk prediction for patients undergoing colorectal surgery

David Chen, PhD[a], Naveed Afzal, PhD[a], Sunghwan Sohn, PhD[a], Elizabeth B. Habermann, PhD[b], James M. Naessens, ScD[b], David W. Larson, MD, MBA[c], Hongfang Liu, PhD[a,*]

[a] Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN
[b] Department of Health Science Research, Mayo Clinic, Rochester, MN
[c] Department of Colorectal Surgery, Mayo Clinic, Rochester, MN

ARTICLE INFO

Article history:
Accepted 23 May 2018
Available online 20 July 2018

ABSTRACT

Background: There is limited consensus regarding risk factors for postoperative bleeding. The objective of this work was to investigate the capability of machine learning techniques in combination with practice-based longitudinal electronic medical record data for identifying potential new risk factors for postoperative bleeding and predicting patients at high risk of postoperative bleeding.
Methods: A retrospective study was conducted for patients who underwent colorectal surgery 1998–2015 at a single tertiary referral center. Various predictors were extracted from electronic medical record. The outcome of interest was the occurrence of postoperative bleeding within 7 days of surgery. Logistic regression and gradient boosting machine models were trained. Area under the receiver operating curve and area under the precision recall curve were used to evaluate the performance to different models.
Results: Of 13,399 cases undergoing colorectal resection, 1,680 (12.5%) experienced postoperative bleeding. A total of 299 variables were evaluated. Logistic regression and gradient boosting machine models returned an area under the receiver operating curve of 0.735 and 0.822 and area under the precision recall curve of 0.287 and 0.423, respectively. In addition to well-known risk factors for postoperative bleeding, nutrition (ranked third), weakness (ranked fifth), patient mobility (ranked sixth), and activity level (ranked eighth) were found to be novel predictors in the gradient boosting machine model based on permutation importance.
Conclusion: The study identified measures of functional capacity of patient as novel predictors of postoperative bleeding. The study found that risk of postoperative bleeding can be assessed, allowing for better use of human resources in addressing this important adverse event after surgery.

Postop Bleeding
AUC 0.82

JAMA Open 2018

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

Abstract

IMPORTANCE Patients with cancer who die soon after starting chemotherapy incur costs of treatment without the benefits. Accurately predicting mortality risk before administering chemotherapy is important, but few patient data–driven tools exist.

OBJECTIVE To create and validate a machine learning model that predicts mortality in a general oncology cohort starting new chemotherapy, using only data available before the first day of treatment.

DESIGN, SETTING, AND PARTICIPANTS This retrospective cohort study of patients at a large academic cancer center from January 1, 2004, through December 31, 2014, determined date of death by linkage to Social Security data. The model was derived using data from 2004 through 2011, and performance was measured on nonoverlapping data from 2012 through 2014. The analysis was conducted from June 1 through August 1, 2017. Participants included 26 946 patients starting 51 774 new chemotherapy regimens.

MAIN OUTCOMES AND MEASURES Thirty-day mortality from the first day of a new chemotherapy regimen. Secondary outcomes included 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.

RESULTS Among the 26 946 patients included in the analysis, mean age was 58.7 years (95% CI, 58.5-58.9 years); 61.1% were female (95% CI, 60.4%-61.9%); and 86.9% were white (95% CI, 86.4%-87.4%). Thirty-day mortality from chemotherapy start was 2.1% (95% CI, 1.9%-2.4%). Among the 9114 patients in the validation set, the most common primary cancers were breast (21.1%; 95% CI, 20.2%-21.9%), colorectal (19.3%; 95% CI, 18.5%-20.2%), and lung (18.0%; 95% CI, 17.2%-18.8%). Model predictions were accurate for all patients (area under the curve [AUC], 0.940; 95% CI, 0.930-0.951). Predictions for patients starting palliative chemotherapy (46.6% of regimens; 95% CI,

Key Points

Question Can a machine learning algorithm applied to electronic health record data predict patients' short-term risk of death at the time that they begin chemotherapy?

Findings In this cohort study of 26 946 patients with cancer starting 51 774 discrete chemotherapy regimens, those at high risk of 30-day mortality were accurately identified across palliative and curative chemotherapy regimens and many types and stages of cancer. The algorithm was more accurate than predictions based on randomized clinical trials or population-based registry data.

Meaning A machine learning algorithm accurately identified individuals at high risk of short-term mortality and may help to guide patient and physician decisions about chemotherapy initiation and advance care planning.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Mortality after chemotherapy
AUC 0.94

A Kaggle Master Explains Gradient Boosting

Ben Gorman | 01.23.2017

This tutorial was originally posted here on Ben's blog, GormAnalysis.

If linear regression was a Toyota Camry, then gradient boosting would be a UH-60 Blackhawk Helicopter. A particular implementation of gradient boosting, XGBoost, is consistently used to win machine learning competitions on Kaggle. Unfortunately many practitioners (including my former self) use it as a black box. It's also been butchered to death by a host of drive-by data scientists' blogs. As such, the purpose of this article is to lay the groundwork for classical gradient boosting, intuitively *and* comprehensively.

Anesth Pain Med 2018;13:248-255
https://doi.org/10.17085/apm.2018.13.3.248
pISSN 1975-5171 · eISSN 2383-7977

인공지능 시대의 마취통증의학 연구

이형철 · 정철우

서울대학교 의과대학 서울대학교병원 마취통증의학교실

앙상블 학습 기법의 대표적인 예로 의사결정나무의 배깅인 랜덤 포레스트(random forest)가 있다[29]. 분류기의 성능을 비교 평가한 연구 결과에 따르면 121개 데이터 셋에 대해 랜덤 포레스트는 비교 대상인 179개 분류기 중 나머지를 압도하는 성능을 보였다[30]. 부스팅 모델의 대표적인 예로는 경사 부스팅(gradient boosting) 모델이 있다. 우수한 예측 성능을 갖는 모형의 상당수가 이 기법을 이용하므로 반드시 자신의 연구에 적용해 보기를 추천한다.

예를 들어 2016년 발표된 패혈증 예측 모델인 InSight의 4시간 전 예측값의 AUROC는 0.74였다[31]. 그러나 2018년에 같은 연구팀이 의사결정나무의 경사 부스팅을 적용한 결과, AUROC는 0.96으로 증가하였다[32]. 2018년 발표된 또 다른 연구는 병원 내 급성 신손상(acute kidney injury, AKI)을 예측하기 위해 의사결정나무의 경사 부스팅을 사용하였으며 AUROC는 0.90였다[33].

감사합니다