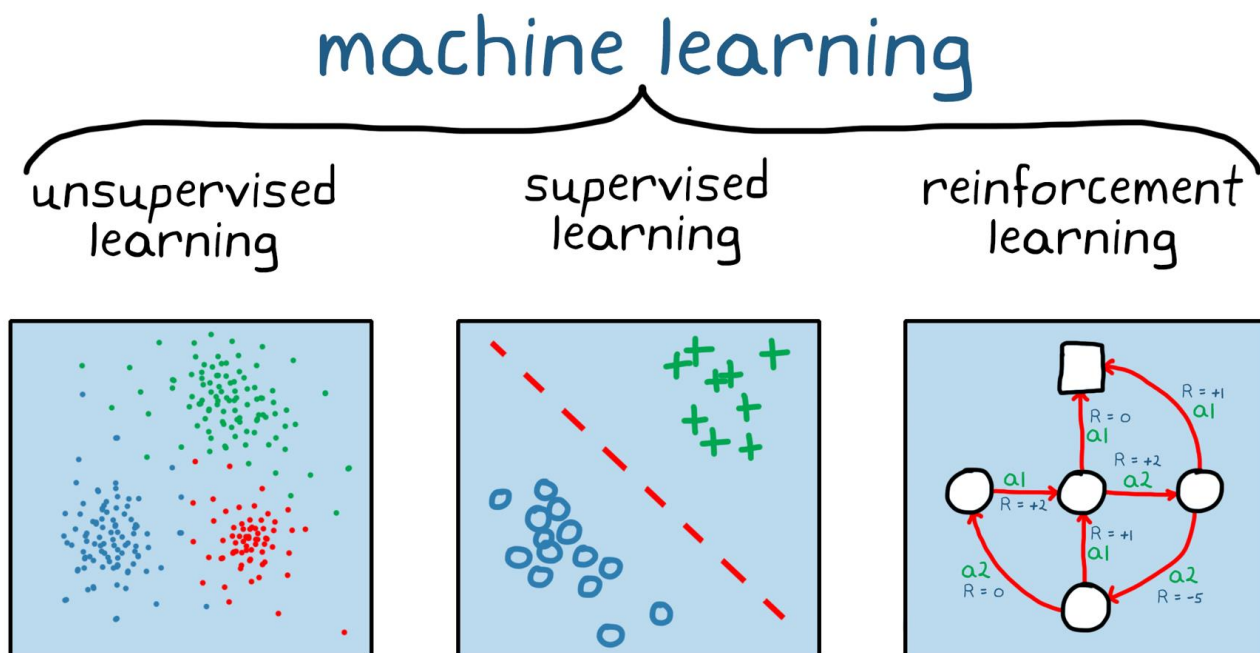


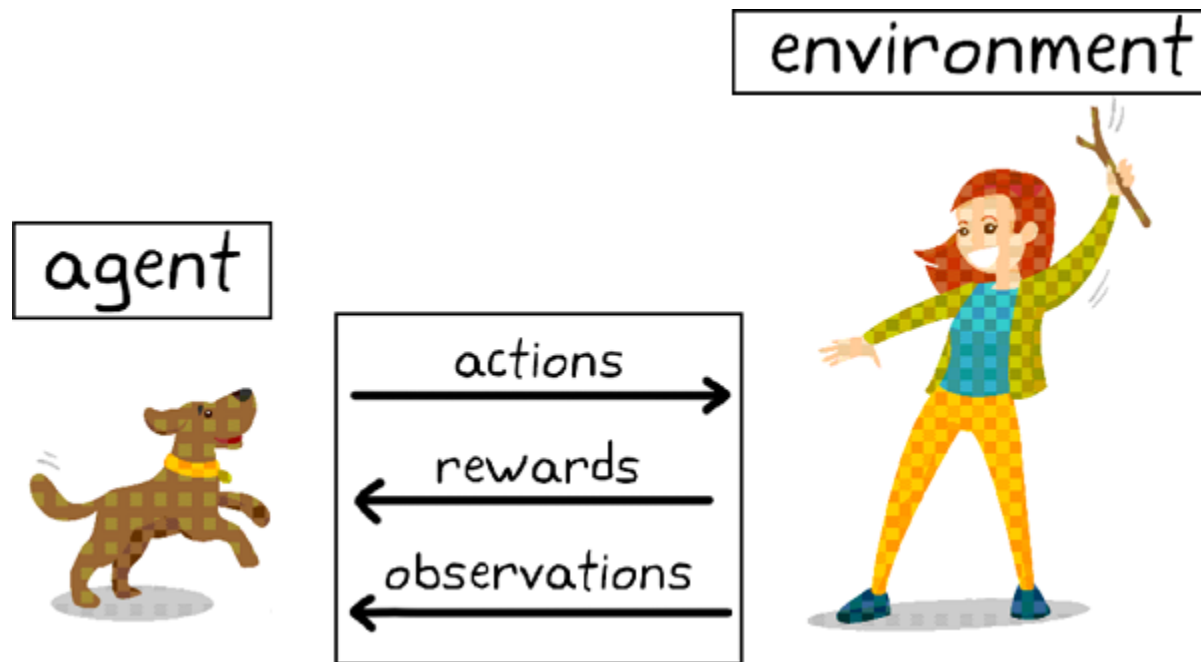
강화학습

# 강화 학습이란? (reinforcement learning; RL)



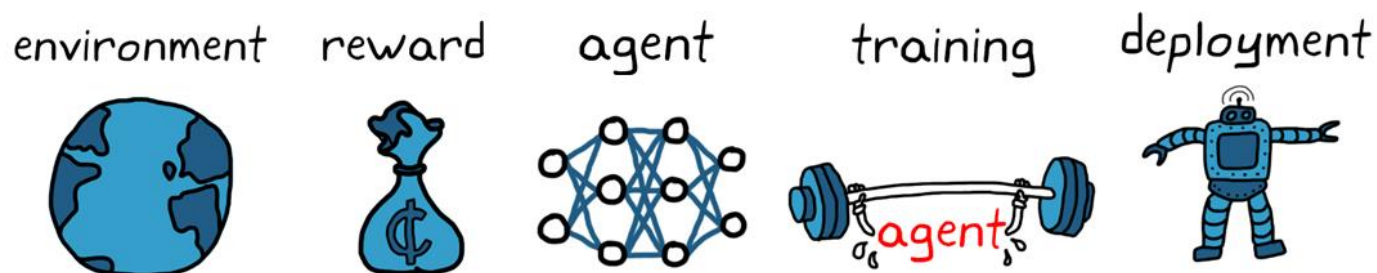
- 머신러닝의 한 분류
- 환경으로부터의 피드백을 기반으로 행위자(agent)의 행동을 분석하고 최적화
- 시행 착오(Trial-and-error)와 지연 보상(delayed reward)을 통해 학습을 하여 목표를 찾아가는 알고리즘
- 최적의 행동양식 또는 정책을 학습하는 것이 목표.

## 강화 학습의 특징



- 비지도 및 지도 머신러닝과 다르게 정적 데이터셋에 의존하지 않고 역동적인 환경에서 동작하여 수집된 경험으로부터 학습
- 지도 및 비지도 머신러닝에서 필요한 훈련 전 데이터 수집, 전처리 및 레이블 지정에 대한 필요성 해소
- 적절한 인센티브가 주어진다면 강화 학습 모델은 인간의 개입 없이 학습을 자체적으로 시작 가능.
- 기존의 신경망들이 label(정답)이 있는 데이터를 통해서 가중치와 편향을 학습하는 것과 비슷하게, 보상(Reward)이라는 개념을 사용하여 가중치와 편향을 학습

# 강화 학습 Workflow



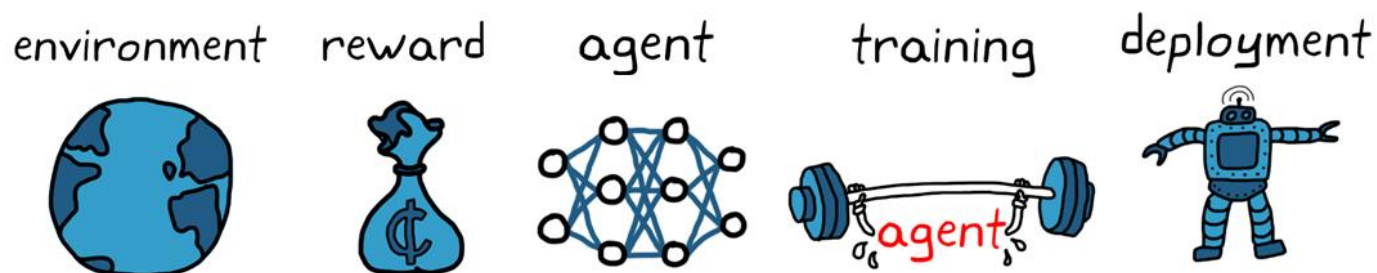
## 1. 환경 생성

먼저 에이전트와 환경 간 인터페이스 등의 강화학습 에이전트가 운영될 환경을 정의해야 함. 시뮬레이션 모델 또는 실제 물리적 시스템일 수 있음. 일반적으로 안전한 실험이 가능한 시뮬레이션 환경을 더 선호

## 2. 보상 정의

에이전트가 성과를 작업 목표와 비교하기 위해 사용할 보상 신호 및 이 신호를 환경으로부터 계산하는 방법을 명시. 보상을 형성하는 것은 까다로운 작업이며 올바르게 설정하기 위해 몇 번의 시도 필요

# 강화 학습 Workflow



## 3. 에이전트 생성

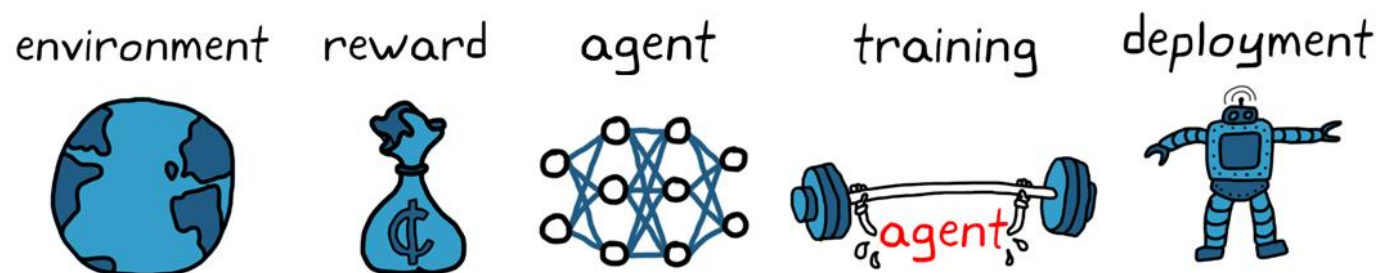
정책과 강화학습 훈련 알고리즘으로 구성된 에이전트 생성.

- 정책을 나타낼 방법 선택(신경망 또는 룩업 테이블 사용 등).
- 적절한 훈련 알고리즘 선택. 일반적으로 최신 강화학습 알고리즘은 상태/행동 공간 및 복잡한 문제에 적합한 신경망 의존.

## 4. 에이전트 훈련 및 검증

- 훈련 옵션(중지 기준 등)을 설정하고 에이전트를 훈련해 정책을 조정.
- 훈련 종료 후 훈련된 정책 검증
- 필요에 따라 보상 신호 및 정책 아키텍처 등의 설계 선택을 다시 검토 및 재훈련

# 강화 학습 Workflow



## 5. 정책 배포

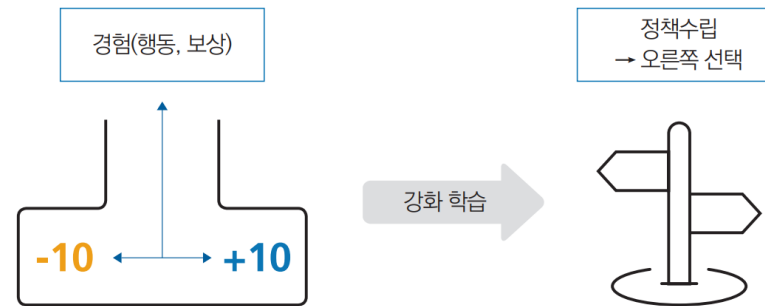
- 훈련된 정책 표현을 코드로 생성하여 배포.

이후, 에이전트가 훈련될 때까지 절차 반복

만약 훈련을 통해 최적의 정책으로 수렴하지 않을 경우, 에이전트를 재훈련하기 전에 다음과 같은 사항을 업데이트 해야함.

- 훈련 설정
- 강화학습 알고리즘 구성
- 정책 표현
- 보강 신호 정의
- 행동 및 관측값 신호
- 환경 동특성

# 강화 학습으로 해결 가능한 문제 유형



강화학습은 결정을 순차적으로 내려야 하는 문제에 적용함.

순차적으로 내려야 하는 문제를 정의하기 위해선 MDP(Markov Decision Process)를 사용

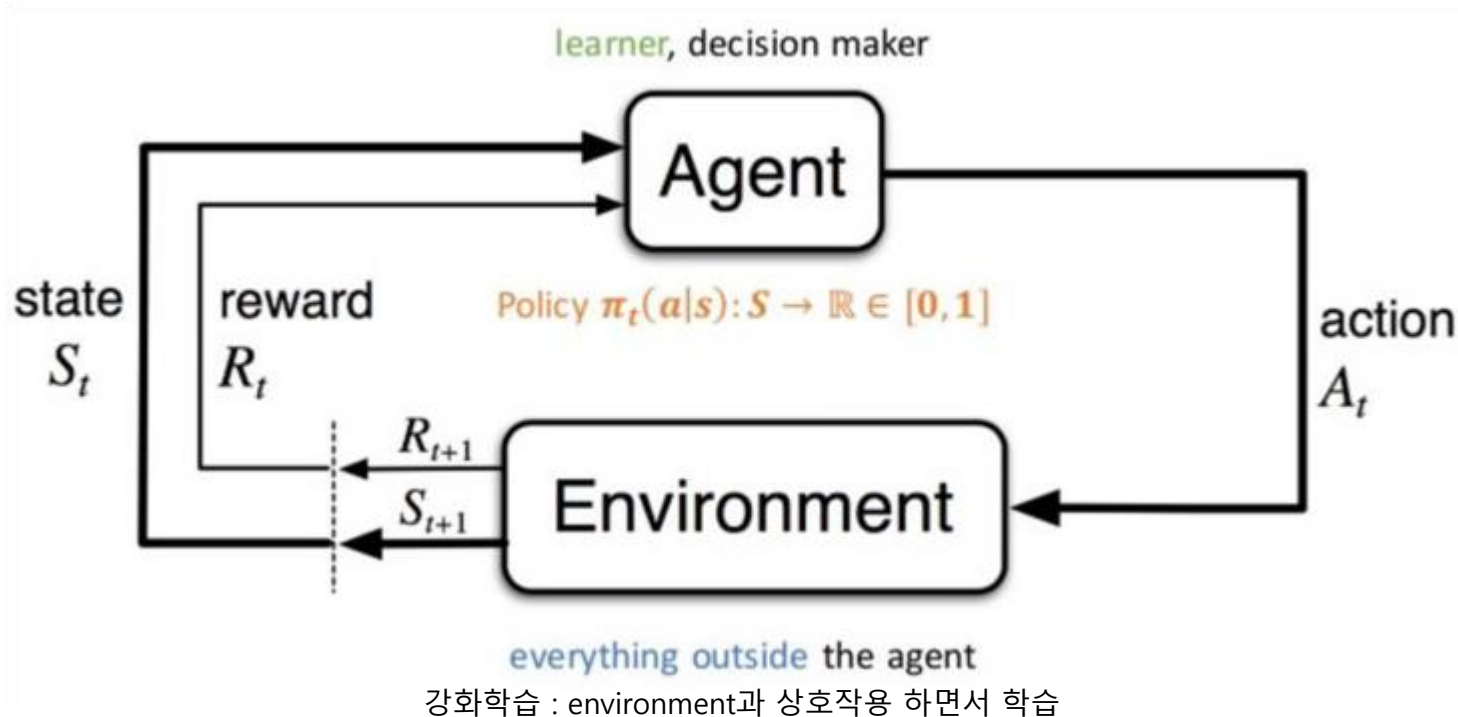
1. 특정 행동에 대한 좋고 나쁨을 평가하는 보상이 주어질 경우
2. 현재의 의사 결정이 이후의 결과값에 영향을 미칠 경우
3. 문제의 구조를 모를 경우
  - 주어진 환경에서 수많은 상호작용을 거쳐 보상 및 결과 정보를 취합

아래와 같은 순서로 문제를 해결해 나간다.

1. 순차적 행동 문제를 MDP로 전환
2. 가치 함수를 벨만 방정식으로 반복적 계산
3. 최적 가치함수와 최적 정책을 찾음.

MP(Markov Process, Chain) : MP는 이산 시간이 진행함에 따라 상태가 확률적으로 변화하는 과정을 의미. 시간 간격이 이산적이고 현재의 상태가 이전의 상태에 영향을 받음. 이것을 기초로 MDP가 정의 됨.

# 강화 학습의 원리\_MDP(Markov Decision Process)



Policy :  $\pi(a | s) = P[A_t=a | S_t=s]$

Optimal policy  $\pi^*$

$$\pi^*(s) = \underset{a \in A(s)}{\operatorname{argmax}} \sum_{s'} P(s'|s, a) v(s')$$

부분 수열(state)의 기대값이 최대가 되는 Policy를 찾는 것이 목표

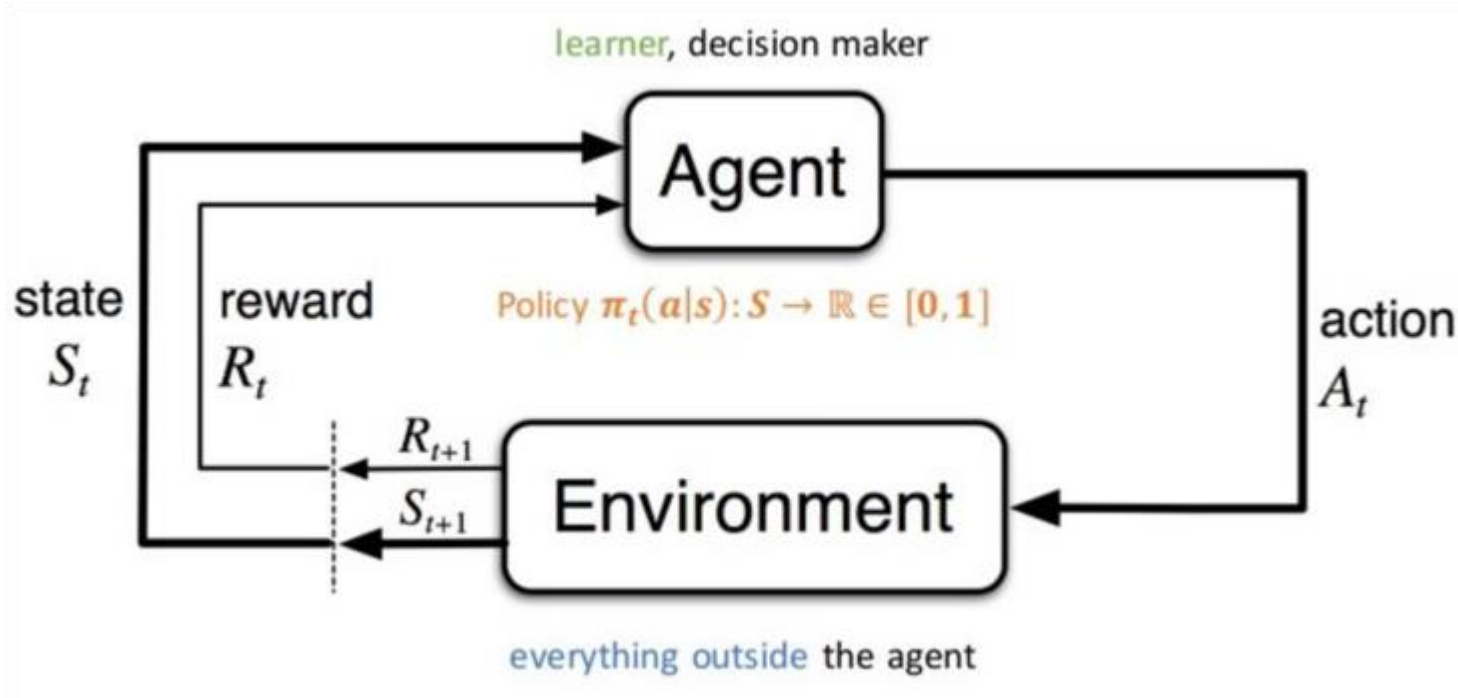
MDP(Markov Decision Process)기반 상태 전이가 현재 상태  $s_t$ 와 입력(행동)  $A_t$ 에 의해 확률적으로 결정되는 모델

학습 방법

1. Agent 가 현재의 상태(state)를 인식하여 어떤 행동(action)을 취함.
2. reward를 받음(positive, negative)
3. 강화 학습의 알고리즘이 Agent가 누적될 포상을 최대화 하려는 행동으로 정의되는 정책을 찾음



## MDP(Markov Decision Process)\_구성요소



- 상태(State) : Agent가 인식하는 자신의 상태 State Set
- 행동(Action) : 특정 State에서 취할 수 있는 행동
- 보상(Reward) : 주어진 상태에서 특정 Action 시 얻는 Reward. Agent가 학습할 수 있는 유일한 정보.
- 정책(Policy) : 모든 상태에 대해 Agent가 행동하는 방식. 순차적 행동 결정 문제(MDP)에서 구해야 할 답.
- 가치 함수(Value Function) : 각 State와 Action이 얼마나 좋은지 추산 및 평가

목표는 최적의 정책(Optimal Policy)을 찾는 것

## 심층 강화학습 (Deep Reinforcement Learning)

강화 학습의 근간을 이루는 점화식을 푸는 가장 대표적인 알고리즘은 동적 계획법(Dynamic Programming)임. 동적 계획법의 연산 복잡도는 Pseudo-Polynomial로써 이는 문제의 상태의 수가 적을 때에는 최적의 해를 주어진 시간내에 연산할 수 있으나, 그렇지 못할 경우에는 매우 많은 연산이 필요함.

Q-Learning과 마르코프 의사 결정 과정 같은 고전적인 강화 학습 알고리즘들은 상태의 수가 커지게 되면 연산 불가능. 따라서 최근 알파고와 같은 복잡한 게임 알고리즘이나 복잡한 통신 시스템에 자동 제어 같은 문제에는 사용 불가.

인공 신경망의 입력과 출력에 각각 상태 정보와 그에 따른 행동 정보를 넣어 학습. 그러면 새로운 상태가 입력되었을 때에 학습에 근거하여 새로운 행동 정보 도출.

이와 같이 인공 신경망 기반의 강화 학습을 딥러닝(Deep Learning)으로 강화학습(Reinforcement Learning)을 수행한다고 하여 심층 강화 학습(Deep Reinforcement Learning)이라고 함.

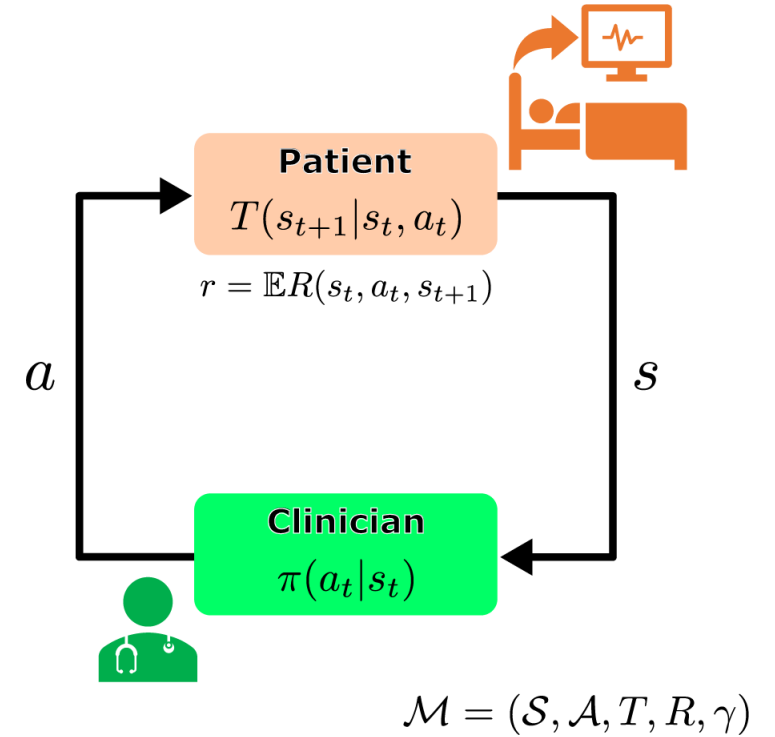
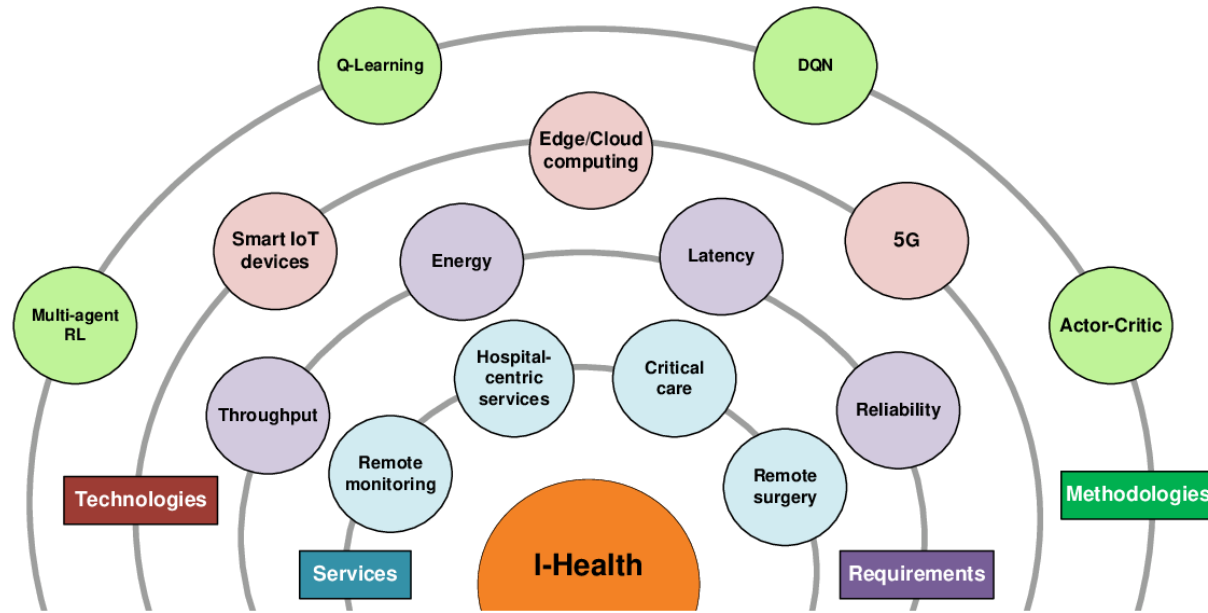
본 알고리즘은 딥러닝 학습 과정의 2단계의 학습 단계에서는 시간이 많이 걸릴 수 있더라도 실제 해당 심층 강화 학습 모델이 사용되는 3단계의 추론에서는 인공 신경망 기본 연산만 하면 되므로 고전적인 강화 학습보다 실제 상황에서는 더욱 빠르게 결론을 도출

## 강화 학습 응용 분야

강화 학습으로 훈련된 심층 신경망은 복잡한 행동을 표현 가능. 이를 통해 기존 방법으로는 해결하기 매우 까다롭거나 어려운 분야에 대안적인 방식으로 접근할 수 있음.

- 고급 제어 : 비선형 시스템을 제어하기는 매우 까다로움. 주로 다양한 동작점에서 시스템을 선형화 하여 해결하는데, 강화학습은 비선형 시스템에 바로 적용 가능
- 자율주행 : 자율주행에서 운전자 대신 카메라 프레임, 라이다 측정값 등 다양한 센서를 동시에 살펴보고 핸들을 어떻게 돌릴지 결정 가능.
- 로봇 공학 : 픽애플레이스 응용 분야에서 로봇 팔로 다양한 사물을 다루는 방법을 학습시키는 등 로봇 파지와 같은 응용 분야에서 사용될 수 있음.
- 스케줄링 : 스케줄링 문제는 신호등 제어, 특정 목표를 위해 공장 현장의 리소스 편성 등 다양한 시나리오가 존재함. 이와 같은 조합 최적화 문제 해결에 있어 좋은 수단
- 보정 : ECU(Electronic Control Unit) 보정과 같이 파라미터의 수동 보정이 필요한 응용 분야에서 활용.

# 의료 분야에서의 강화학습



의료에서 순차적 의사 결정: 임상 의사는 환자의 State 관찰(s), 적절한 치료 방법 선택 (a)한 뒤 모니터링. 프로세스 반복. 환자의 상태가 전환될 때마다 (T, 확률) 보상이 주어짐 (R).

정보를 바탕으로 행한 행위의 다음 상황에 대해 스스로 추정치를 계산함으로써 실제와 예상치를 줄여나가는 강화학습은 동적이고 불확실한 환경에서도 최적의 상황으로 수렴할 수 있도록 도와준다.

- 의료 현장에서 어떤 환자의 데이터와 파악할 질병 상황을 바탕으로, 어떤 치료를 순차적으로 적용해야 완치라는 단계로 나아갈도록 조절할 수 있는지 전략을 세우는데 활용 가능

# 의료 분야에서의 강화학습

## 1. Precision Medicine

- 각 환자의 개별 특성을 파악하여 개별 치료가 가능하게 도와줌. 환자의 병의 진행 정도나 상태에 따라 약의 투여량을 학습

## 2. Dynamic Treatment Regime(DTR)

- DTR은 단계마다 하나씩, 변해가는 치료 및 상태를 기반으로 환자에게 개별적으로 알맞은 치료 방법을 학습 및 지시. 개인화된 자료(진단 테스트 결과, 유전 데이터, 인구통계학적 정보 등)와 환자의 상태를 고려하여 실시간으로 맞춤 치료를 제공

## 3. Medical Imaging

- 의료 임상 분석을 돕기 위해 인체 또는 일부 내부 조직의 시각적 표현을 생성하거나 병변을 분류, detection하는 방법을 학습. MRI, 방사선 사진, 탄성조영술, 단층촬영, 초음파, 광음향영상 등에 사용.

## 4. Diagnostic systems

- 환자의 증상과 상태를 통해 의사의 결정을 지원.

## 5. Dialogue systems, chat-bots and advanced interfaces

- 인간 간의 채팅 특성을 모방하여 환자를 케어하거나 인지 장애 진단. 또한 온라인 증상 검사기에 쓰여 질병을 예측하거나 환자에 증상에 따라 치료법을 제안.

## 6. Control systems

- 보조 장치에 이용하여 환자의 팔 움직임을 돕거나, 환자의 상태를 파악하여 인슐린 주입을 돕는 역할

## 7. Rehabilitation

- 가상 재활 시스템에 사용되며 난이도를 조절하거나 치료사 보조 역할 수행,

## 8. Health Management Systems

- 예약, 자원 및 물품 관리, 비용 절감 및 품질 개선을 포함하여 의료 서비스 조직과 관련된 역할이 포함됨.

# 1. Precision Medicine

## 패혈증 환자에 대한 승압제 및 수액 치료 전략 수립 사례

**naturemedicine**

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature medicine](#) > [articles](#) > article

Article | [Published: 22 October 2018](#)

### **The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care**

[Matthieu Komorowski](#), [Leo A. Celi](#), [Omar Badawi](#), [Anthony C. Gordon](#)  & [A. Aldo Faisal](#) 

[Nature Medicine](#) **24**, 1716–1720 (2018) | [Cite this article](#)

**43k** Accesses | **377** Citations | **648** Altmetric | [Metrics](#)

- 중환자실에서 치료했던 데이터를 기반으로 가상의 임상 환경을 마련해 두고, 그 안에서 시행착오를 경험하여 적절한 승압제와 수액 치료를 할 수 있도록 인공지능 모델을 학습시킴
- 그 결과 실제 의료진의 치료 정책보다 인공지능의 치료 정책이 저 많은 환자를 살릴 수 있을 것이라고 평가됨

<https://www.nature.com/articles/s41591-018-0213-5>

# Data

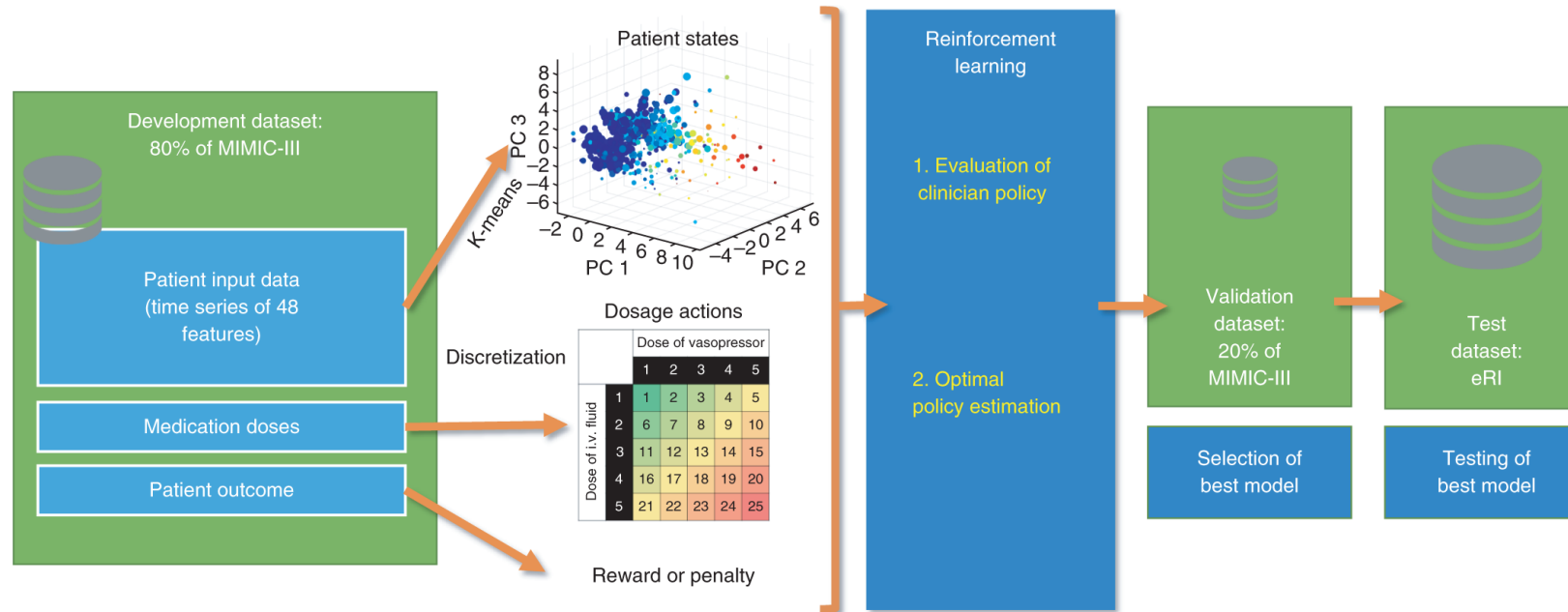
- \*MIMIC-III는 모델 개발, \*eRI는 모델 검증에 사용
- 두 데이터 세트에서 패혈증 기준을 충족하는 성인 환자 포함
- 인구 통계, Elixhauser 병전 상태, 활력 징후, 실험실 값, 받은 체액 및 승압제, 체액 균형 을 포함한 48개 변수 세트 추출
- 최종 예측 목표인 정맥주사(iv) 수액과 혈관수축제 용량의 조합은 25가지로 구분

Discretized action	IV fluids (mL/4 hours)		Vasopressors (mcg/kg/min)	
	Range	Median dose	Range	Median dose
1	0	0	0	0
2	]0-50]	30	]0-0.08]	0.04
3	]50-180]	85	]0.08-0.22]	0.13
4	]180-530]	320	]0.22-0.45]	0.27
5	>530	946	>0.45	0.68

- (i) MIMIC-III : 2001~2012년까지 보스턴 교육 병원의 6개 중환자실에 입원한 61,532명의 오픈 데이터베이스
- (ii) Philips eRI : 2003~2016년까지 미국 전역의 459개 ICU에서 330만 건 이상의 입원 기록

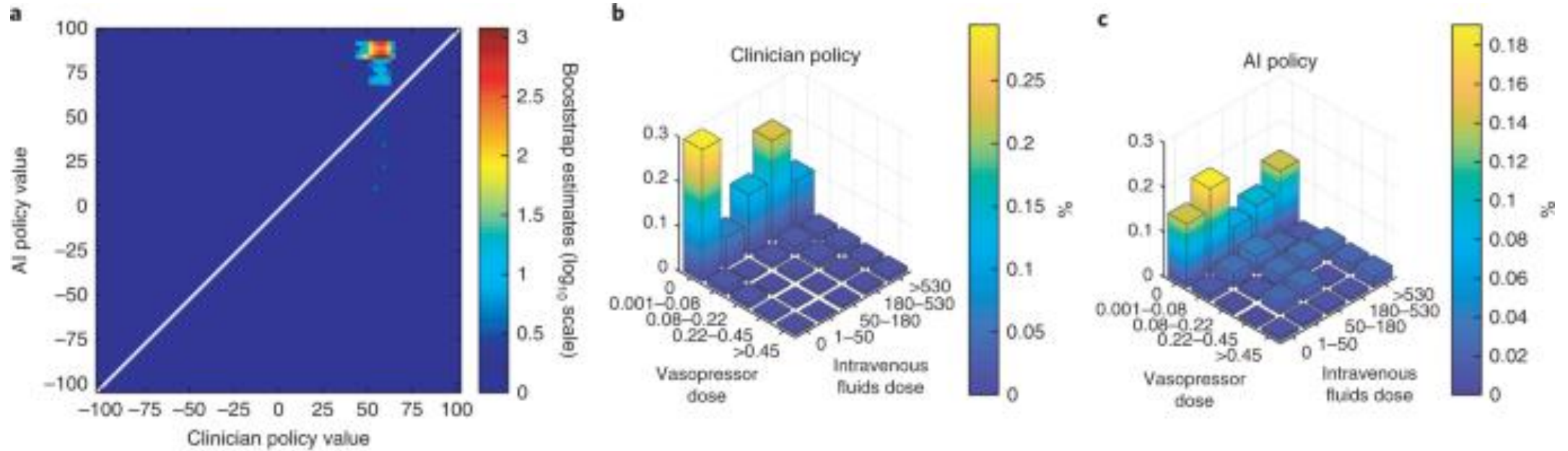
# Model

- 정맥 수액 및 승압제의 모든 처방을 관찰하고 각 치료 옵션의 평균값을 계산하여 Q 함수의 시간차 학습(TD-learning)을 사용하여 임상치의 실제 행동(정책) 평가를 수행했습니다.
- TD 학습의 장점은 MDP(모델 프리)에 대한 지식이 필요하지 않고 샘플 궤적에서 간단히 학습할 수 있다는 것입니다.

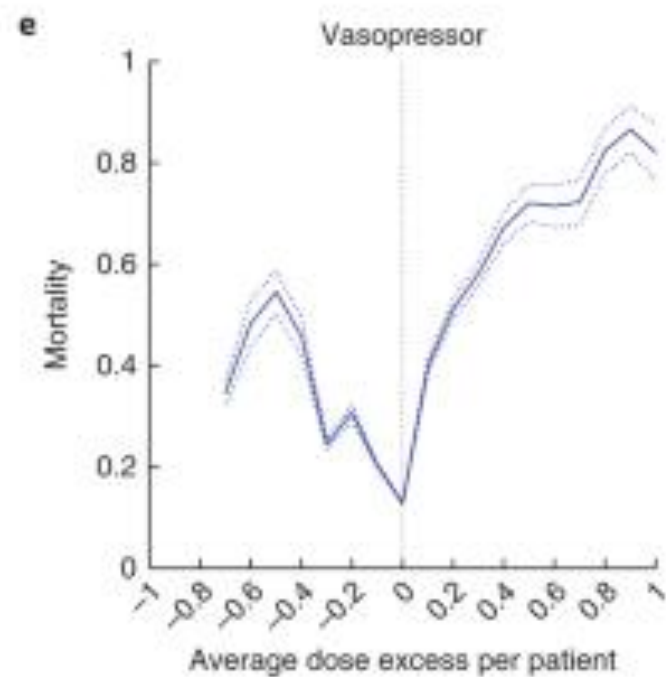
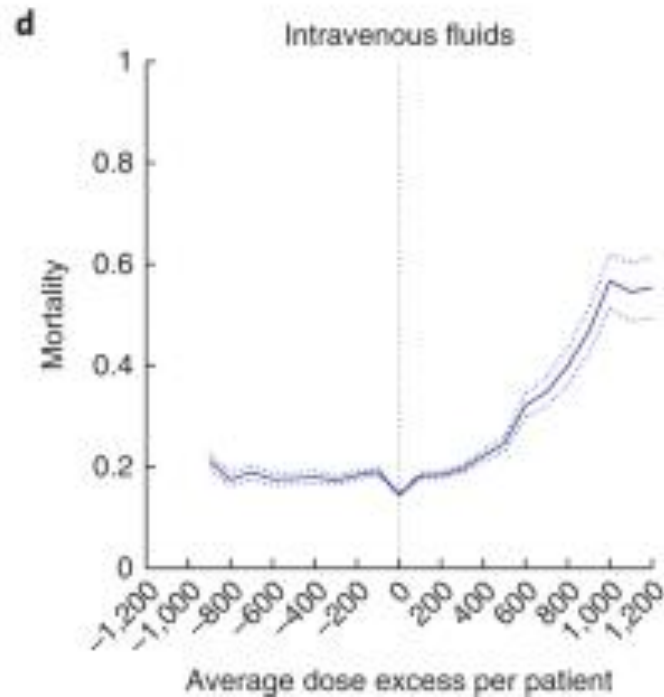




# Result



평균적으로 환자들은 AI 정책에서 권장하는 것보다 더 많은 정맥 수액(b)과 더 적은 양의 혈압 상승제(c)를 투여 받았습니다.



그래프는 용량 초과, 정맥 수액(왼쪽) 및 승압제(오른쪽)에 대해 환자당 모든 시점에 대해 평균을 낸 제공된 용량과 제안된 용량 간의 차이를 나타냅니다.

두 그림에서 가장 작은 용량 차이가 최고의 생존율을 보였습니다.  
받은 복용량이 제안된 복용량에서 멀어질수록 생존율은 더 낮아졌습니다.

## 2. Dynamic Treatment Regime(DTR)

폐암에서 자동화된 방사선 적응을 위한 심층 강화 학습

# MEDICAL PHYSICS

The International Journal of Medical Physics Research and Practice

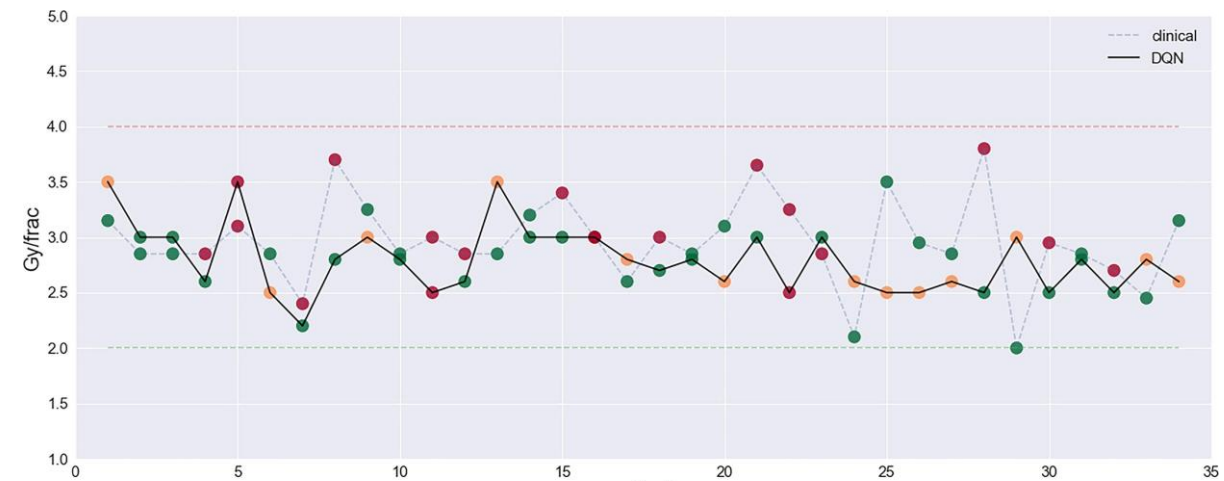
Research Article | [Free Access](#)

### Deep reinforcement learning for automated radiation adaptation in lung cancer

Huan-Hsin Tseng, Yi Luo, Sunan Cui, Jen-Tzung Chien, Randall K. Ten Haken, Issam El Naqa

First published: 16 October 2017 | <https://doi.org/10.1002/mp.12625> | Citations: 95

- 방사선 폐렴 등급 2(RP2)을 최소한의 종양 국소 제어를 통해 최대의 효과를 만드는 것이 목표
- GAN과 사용하여 표본 크기로부터 강화학습에 필요한 특성을 학습, DNN을 사용하여 환자의 치료 과정 상태를 학습, 강화학습(DQN Deep Q-network)를 통해 치료 환경에서 최적의 용량을 선택을 학습.
- 결과적으로, 이 기준은 임상이가 19개의 좋은 결정과 15개의 나쁜 결정을 내린 반면 DQN은 표 2 에 정리된 것처럼 17개의 좋은 결정, 4개의 나쁜 결정 및 13개의 잠재적으로 좋은 결정을 내렸음을 시사합니다. 결과는 "좋은" 범주에서 임상가와 DQN 사이에 유의미한 차이가 없는 반면 "나쁨" 및 "잠재적으로 양호" 범주에는 차이가 있어 이러한 불확실한 상황에서 DQN에 유리할 수 있음을 보여줌.
- 강화학습에 의한 방사선 용량과 임상가가 선택한 결과가 유사하게 나옴. 하지만 이 프레임워크를 신뢰할 수 있게 만들려면 더 큰 데이터 세트에 대한 추가 검증 필요



RMSE = 0.5 Gy인 DQN(검은색 실선) 대 임상 결정(파란색 점선)에 의해 제공된 자동 용량 결정. 좋은(녹색 점), 나쁜(빨간색 점) 및 잠재적으로 좋은 결정(주황색 점)의 평가는 표 2

Table 2. Summary for the evaluation on clinicians' and the DQN decisions

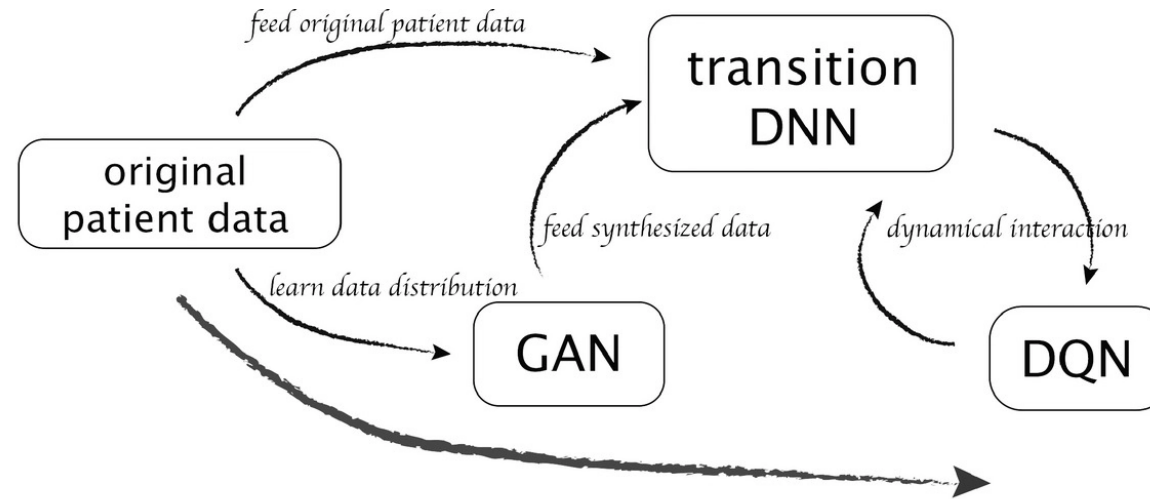
Summary	Good	Bad	Potentially good
clinicians	19 (55.9%)	15 (44.1%)	0
DQN	17 (50%)	4 (11.8%)	13 (38.2%)

# Data

- 반응 기반 용량 적응의 결정 지원을 위해 114명 NSCLC 환자의 과거 치료 계획 사용
- RAE를 모델링하기 위해 표에 설명된 특성을 가진 14개 선택

예측 변수	생물학적/임상적 특성	참고문헌
IL4	Th2 사이토카인. 항체 생성, 조혈 및 염증을 조절합니다. 순진한 헬퍼 T 세포의 Th2 세포로의 분화를 촉진합니다. Th1 세포 생산 감소	26 - 28
IL15	Th1 사이토카인. 자연사(NK) 세포의 활성화 및 세포독성을 유도합니다. 대식세포를 활성화합니다. T 및 B 림프구 및 NK 세포의 증식 및 생존 촉진	26 - 28
GLSZM.GLN	GLSZM(Gray-Level Size Zone Matrix)의 GLN(Gray-Level Nonuniformity) 기능은 다음과 같이 정의됩니다. $\sum_{i=1}^{N_g} (\sum_{j=1}^{N_r} P(i, j))^2 / \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i, j)$ .	표기법은 Ref. 29
GLRLM.RLN	다음으로 정의된 GLRLM(그레이 레벨 런 길이 행렬)의 RLN(런 길이 비균일성) 기능 $\sum_{j=1}^{N_g} (\sum_{i=1}^{N_r} P(i, j))^2 / \sum_{j=1}^{N_g} \sum_{i=1}^{N_r} P(i, j)$ .	표기법은 Ref. 29
MCP1	케모카인은 지방세포에서 발현되고 분비됩니다. <i>MCP-1</i> 의 지방세포 발현은 <i>TNF-α</i> 에 의해 증가	30 , 31
TGFβ1_ _	Th2 사이토카인. 면역 체계의 조절에 중요한 역할. 림프구, 대식세포 및 수지상 세포를 포함한 모든 백혈구 계통에서 생성	32
폐/종양 gEUD( $\alpha / \beta$ =4 Gy ,10 Gy resp)	EQD2 용량 분포에서 변환된 폐/종양의 일반화 등가 균일 용량: $EQD_2 = N_{\text{frac}} \times d \times \left( \frac{d + \alpha / \beta}{2 + \alpha / \beta} \right)$	33 , 34
MTV	PET 영상의 대사 종양 부피	—

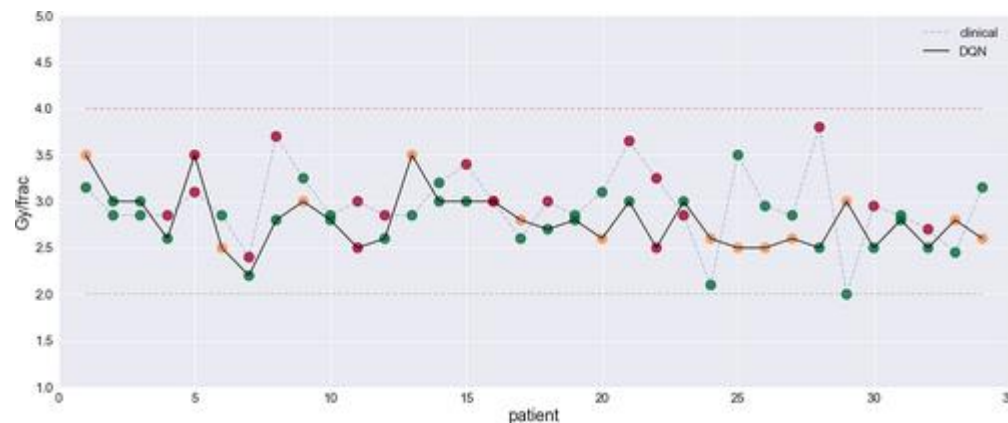
# Model



**(b)** A 3-component DNN solution is proposed to (1) generate synthetic data through GAN to (2) model the radiotherapy environment by the transition DNN that is used by (3) the DQN to make optimal decisions for adaptation of dose.

- GAN과 사용하여 표본 크기로부터 강화 학습에 필요한 특성을 학습
- DNN을 사용하여 환자의 치료 과정 상태를 학습
- 강화 학습(DQN)를 통해 치료 환경에서 최적의 용량을 선택을 학습

# Result



요약	좋은	나쁜	잠재적으로 좋음
임상의	19(55.9%)	15 (44.1%)	0
DQN	17 (50%)	4 (11.8%)	13 (38.2%)

"좋은"과 "나쁨"을 녹색과 빨간색 점 사용하여 임상의와 DQN이 내린 결정의 품질 확인 가능

"좋은" 범주에서 임상의와 DQN 사이에 유의미한 차이가 없는 반면 "나쁨" 및

"잠재적으로 양호" 범주에는 차이가 있어 이러한 불확실한 상황에서 DQN에 유리할 수 있음



### 3. Medical Imaging

#### 강화학습을 통한 3D 의료 이미지 Segmentation 사례



This CVPR 2020 paper is the Open Access version, provided by the Computer Vision Foundation.  
Except for this watermark, it is identical to the accepted version;  
the final published version of the proceedings is available on IEEE Xplore.

#### Iteratively-Refined Interactive 3D Medical Image Segmentation with Multi-Agent Reinforcement Learning

Xuan Liao<sup>1</sup>, Wenhao Li<sup>\*2</sup>, Qisen Xu<sup>\*2</sup>, Xiangfeng Wang<sup>2</sup>✉,  
Bo Jin<sup>2</sup>✉, Xiaoyun Zhang<sup>1</sup>, Yanfeng Wang<sup>1</sup>, and Ya Zhang<sup>1</sup>✉

<sup>1</sup> Cooperative Medianet Innovation Center, Shanghai Jiao Tong University

<sup>2</sup> Multi-agent Artificial Intelligence Laboratory, East China Normal University

{liaoxuan, xiaoyun.zhang, wangyanfeng, ya\_zhang} @sjtu.edu.cn, {51164500105@stu,  
51184501067@stu, xfwang@sei, bjin@cs} .ecnu.edu.cn

Update \ Initial	BG	V-Net	HighRes3DNet	DeepIGeoS(P-Net)
Initial	0	77.15	75.39	82.16
Min-cut	27.46	80.69	77.05	84.08
DeepIGeoS(R-Net)	82.97	85.80	85.72	84.83
InterCNN	85.17	85.56	87.29	86.54
IteR-MRL	<b>86.14</b>	<b>88.53</b>	<b>87.43</b>	<b>87.50</b>

Table 1. Combination with different initial methods

Step	0	1	2	3	4	5
Clicks	0	5	10	15	20	25
Min-Cut	77.15	79.52 (+2.37)	79.97 (+0.45)	80.22 (+0.25)	80.46 (+0.24)	80.69 (+0.23)
DeepIGeoS(R-Net)	77.15	<b>85.62</b> (+8.47)	85.74 (+0.12)	85.73 (-0.01)	85.75 (+0.02)	85.80 (+0.05)
InterCNN	77.15	83.19 (+6.04)	84.39 (+1.20)	85.16 (+0.77)	85.52 (+0.36)	85.56 (+0.04)
IteR-MRL	77.15	84.35 (+7.20)	<b>86.78</b> (+2.43)	<b>87.61</b> (+0.83)	<b>88.18</b> (+0.57)	<b>88.53</b> (+0.35)

Table 2. Performance improvement in one interactive sequence

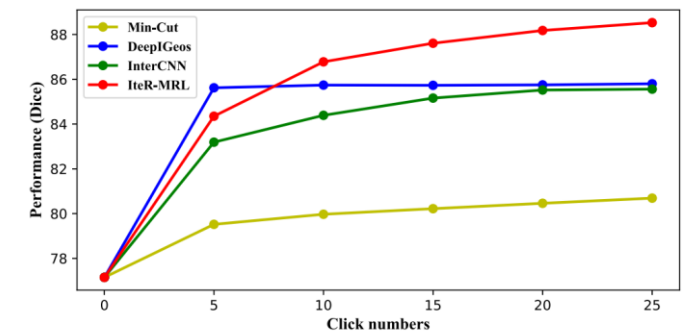


Figure 4. Performance improvement shown by curves

- 의료 3D 영상 segmentation은 좋은 품질을 위해서는 수동 주석이 필요하고 의사의 경험에 의해 성능이 좌우됨. 또한 기존 3D segmentation(CNN)는 임상 용도로는 부족한 상태임.
- 이를 해결하기 위해 Markov decision process (MDP)로 강화학습하여 기존의 CNN 방법보다 수렴이 빠르고 좋은 성능을 냄.

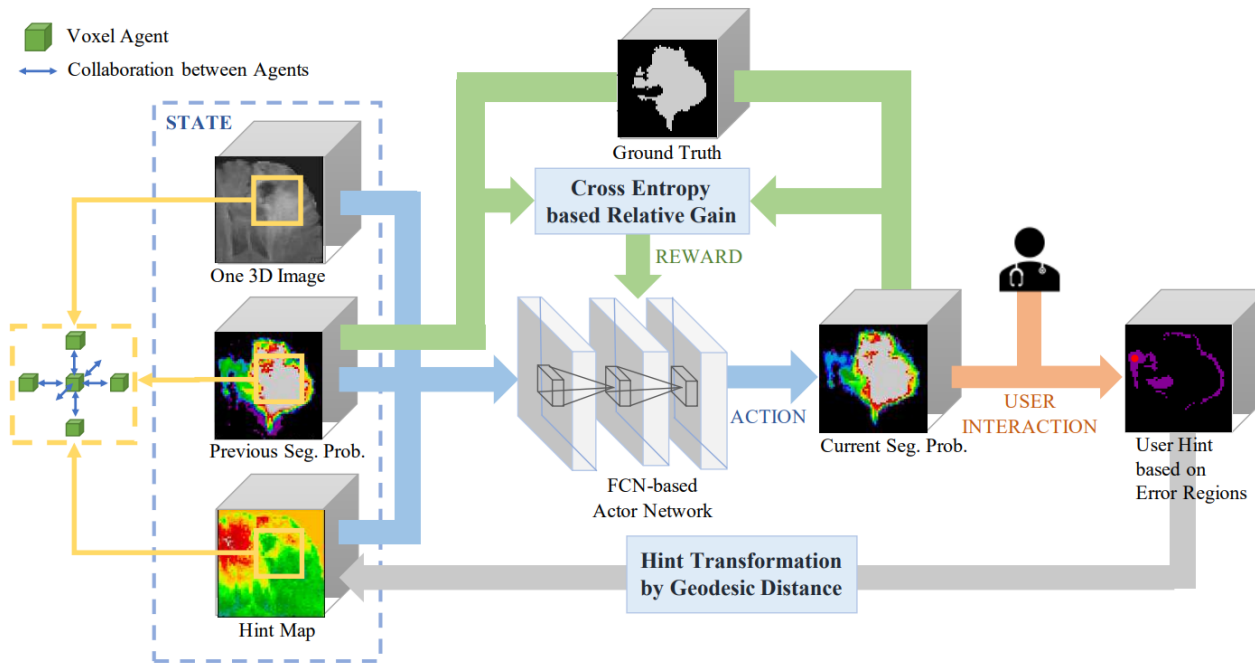


Figure 2. Overview of Iteratively-Refined interactive 3D medical image segmentation algorithm based on MARL (IteR-MRL). At each refinement step, the state containing image, previous segmentation probability and the hint map is fed into the actor network, then the actor network produces current segmentation probability derived by its output actions. Next, the user gives back hint clicks (the red point) based on error regions and new hint map is generated by hint transformation. At every step, the reward is determined by the relative gain between previous and current segmentation cross entropy. Voxels are regarded as agents who collaborate with each other in our method.

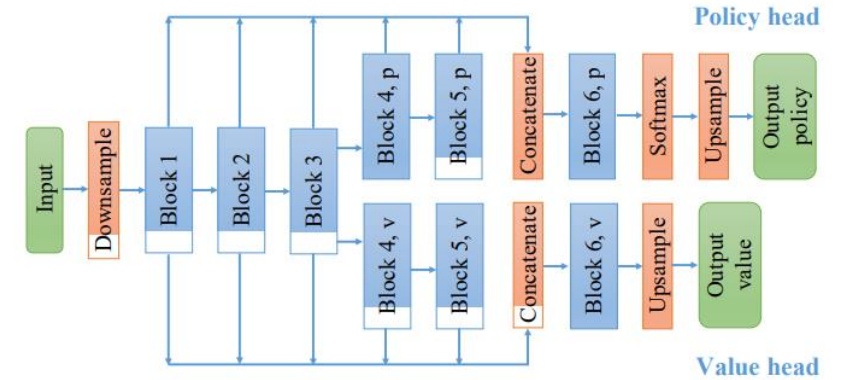


Figure 3. The network architecture for IteR-MRL. The policy and value heads share the low-level features and extract their own high-level features.

- 원본 이미지, 이전 모델을 통과하여 나온 segmentation 이미지, 의사로부터 받은 hint map을 모델 인풋으로 사용함. 이를 통해 Segmentation 주석 output를 만듦.
- 사용자가 오류 이미지에 빨간 점은 표시에 hint map를 만들어 다시 input으로 활용
- 현재 정책에서 나온 segmentation 이미지와 전 정책에서 만들어진 이미지의 cross entropy를 계산(이미지, 힌트맵, 이전 segmentation 확률)하여 보상을 줌.



## 4. Diagnostic systems

### RL 기반 II형 당뇨병 위험 예측

Home / Proceedings / ICIEV-&-ICIVPR / ICIEV-&-ICIVPR 2020

2020 Joint 9th International Conference on Informatics, Electronics & Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision & Pattern Recognition (icIVPR)

### Forecasting the Risk of Type II Diabetes using Reinforcement Learning

Year: 2020, Pages: 1-6

DOI Bookmark: 10.1109/ICIEVICIVPR48672.2020.9306653

#### Authors

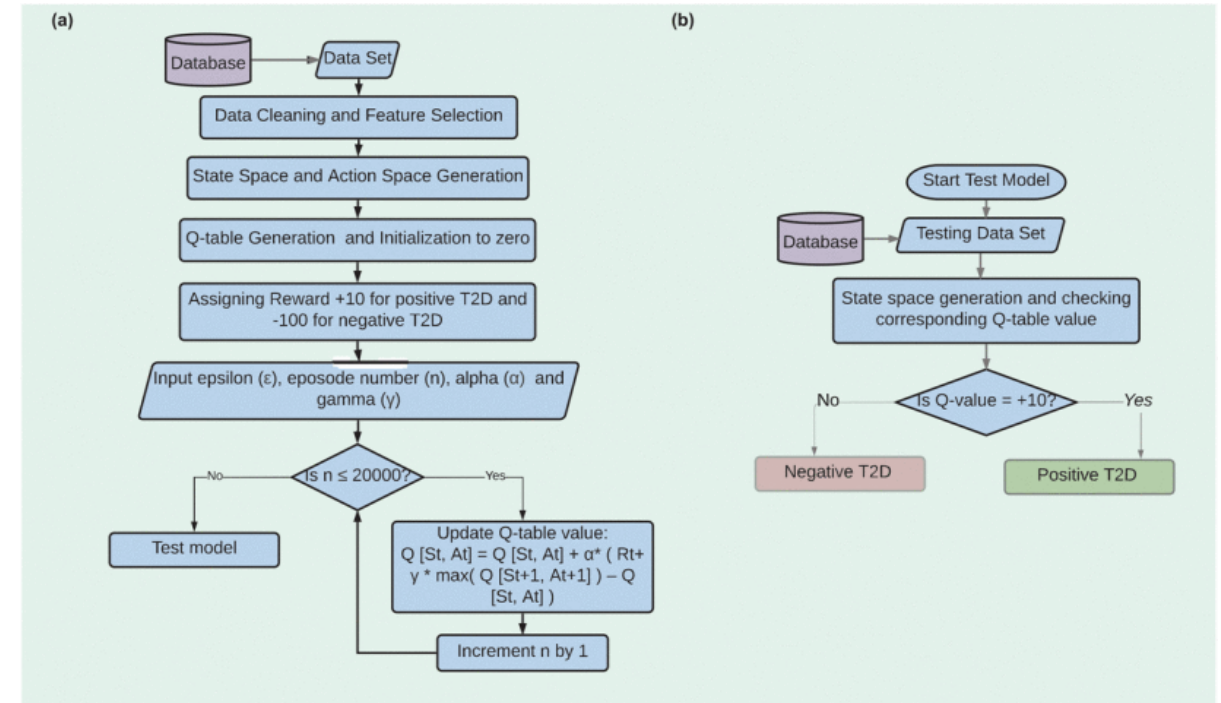
Most. Fatematuz Zohora, Institute of Information Technology, Jahangirnagar University, Dhaka, Bangladesh, 1340

Marzia Hoque Tania, Institute of Biomedical Engineering, University of Oxford, Department of Engineering Science, UK

M Shamim Kaiser, Institute of Information Technology, Jahangirnagar University, Dhaka, Bangladesh, 1340

Mufti Mahmud, Nottingham Trent University, Dept. of Computing & Technology, Nottingham, UK, NG118NS

- 2형 당뇨병(T2D)을 EHR(Electronic Health Records)[체질량 지수, 포도당 수준 및 대상자의 연령]을 통해 식별
- KNN과 DT를 RL 모델과 비교하여 예측의 성능이 증가하는 것을 확인
- 중요한 생활 습관 정보만을 가지고 T2D를 식별하여 의료 전문가의 치료 효율을 줄이는데 도움이 될 수 있고, T2D 발병 역치 수준에 매우 가까운 환자의 당뇨병 전증 상태를 식별할 수 있었음.
- 추후 다른 요인을 더한다면 더 높은 성능을 보일 것이라고 예상함.



## 6. Control systems

### RL 기반 개인화 인슐린 용량 제안 사례


Journals & Magazines > IEEE Journal of Biomedical an... > Volume: 23 Issue: 6 ?

#### A Dual Mode Adaptive Basal-Bolus Advisor Based on Reinforcement Learning

Publisher: IEEE

Cite This

PDF

Qingnan Sun ; Marko V. Jankovic; João Budzinski; Brett Moore; Peter Diem; Christoph Stettler; Stavroula G. Mougi... [All Authors](#)

17

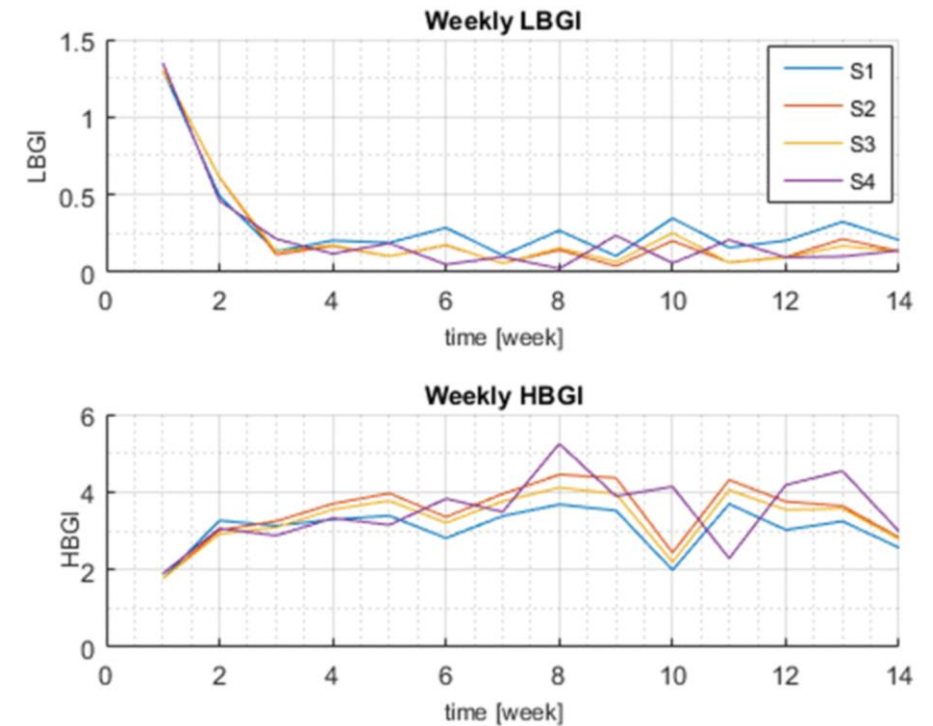
Paper

Citations

634

Full

Text Views



- RL 를 통해 ABBA(adaptive basal-bolus algorithm) 를 제안. 전날 환자의 포도당을 기반으로 인슐린 용량에 대한 개인화된 투여량 제안하여 저혈당 위험 최소화를 목표
- SMBG(self-monitoring of blood glucose)와 CGM(continuous glucose monitoring) 데이터를 기반으로 학습되며 환자 개인 설정CIR(correction factor, insulin-to-carbohydrate ratio), BR(basal rate)를 갱신하여 개인화된 인슐린 투여량을 제안
- SMBG(self-monitoring of blood glucose)와 CGM(continuous glucose monitoring) 데이터를 기반으로 학습되며 환자 개인 설정CIR(correction factor, insulin-to-carbohydrate ratio), BR(basal rate)를 갱신하여 개인화된 인슐린 투여량을 제안
- 표준치료했을 때와(실험 실시 이전), 13주 동안의 ABBA 를 이용했을 때의 차이를 비교해보니 저혈당위험지수(LBGI)는 감소하고 고혈당위험지수(HBGI)는 최소 범위(<https://www.researchgate.org/publication/3579186>)를 유지함으로써, 가능성을 보여주었다.

# 8. Health Management Systems

RL 기반 환자 예약을 통한 응급실 효율성 향상



Healthcare (Basel). 2020 Jun; 8(2): 77.

Published online 2020 Mar 27. doi: [10.3390/healthcare8020077](https://doi.org/10.3390/healthcare8020077)

PMCID: PMC7349722

PMID: [32230962](https://pubmed.ncbi.nlm.nih.gov/32230962/)

## Improving Emergency Department Efficiency by Patient Scheduling Using Deep Reinforcement Learning

Seunghoon Lee and Young Hoon Lee\*

▶ Author information ▶ Article notes ▶ Copyright and License information ▶ Disclaimer

그림 3. 기존 디스패치 규칙과 RL을 사용하는 규칙 간의 비교.

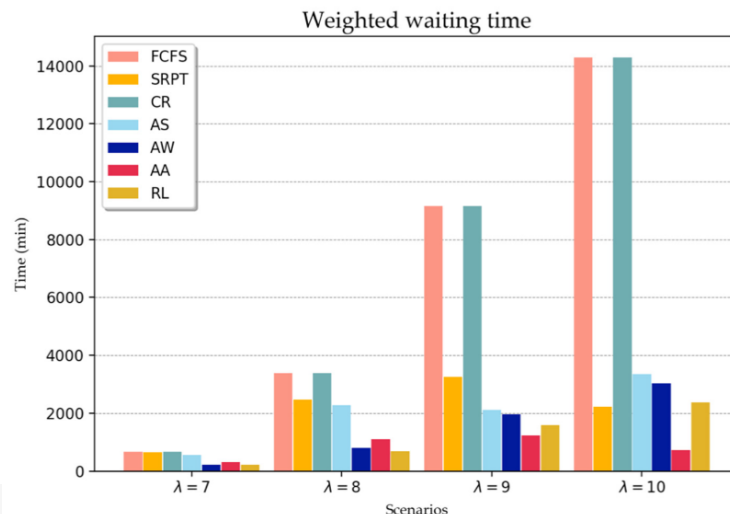


Table 5. Dispatching rules.

Dispatching Rule	Description
First Come, First Served (FCFS)	Select the patient arriving at the earliest time in the ED
Shortest Remaining Processing Time (SRPT)	Select a patient having the shortest remaining average process time
Critical Ratio (CR)	Select a patient by dividing the actual time remaining for a particular treatment by the estimated time required for the entire treatment process
Acuity and SRPT (AS)	Select a patient based on the linear combination of two weighted variables: weighted acuity and SRPT
Acuity and Waiting Time (AW)	Select a patient based on the linear combination of two weighted variables: weighted acuity and weighted waiting time
Acuity and Arrival Time (AA)	Select a patient considering the high acuity level and earliest arrival time as primary and secondary factors, respectively

- 수학적 모델과 Markov Decision Process로 공식화하고, Q-network(DQN)으로 환자를 예약하기 위한 최적의 정책을 결정
- ATS(Australian Triage scale)과 CTAS(Canadian triage and acuity scale)로 환자의 상태를 1~5단계로 구분
- 비교할 dispatching rules 과 개발한 RL 를 비교
- AA는 그래프상 RL보다 적은 시간을 대기하게 하는것 같지만, 높은 위험 환자를 먼저 퇴원시켜 적은 위험 환자는 아예 퇴원할 수 없는 것으로 나타남. 반면에 RL은 모든 중증도에서 적절한 수의 환자가 퇴원함.
- ED에서의 의사결정이 Deep RL을 사용하여 접근 가능하다는 것을 보여줌.

# 단기간 동안 신약 후보 물질 디자인 가능 사례

Brief Communication

<https://doi.org/10.1038/s41587-019-0224-x>

nature  
biotechnology

## Deep learning enables rapid identification of potent DDR1 kinase inhibitors

Alex Zhavoronkov<sup>1\*</sup>, Yan A. Ivanenkov<sup>1</sup>, Alex Aliper<sup>1</sup>, Mark S. Veselov<sup>1</sup>, Vladimir A. Aladinskiy<sup>1</sup>, Anastasiya V. Aladinskaya<sup>1</sup>, Victor A. Terentiev<sup>1</sup>, Daniil A. Polykovskiy<sup>1</sup>, Maksim D. Kuznetsov<sup>1</sup>, Arip Asadulaev<sup>1</sup>, Yury Volkov<sup>1</sup>, Artem Zholus<sup>1</sup>, Rim R. Shayakhmetov<sup>1</sup>, Alexander Zhebrak<sup>1</sup>, Lidiya I. Minaeva<sup>1</sup>, Bogdan A. Zagribelnyy<sup>1</sup>, Lennart H. Lee<sup>2</sup>, Richard Soll<sup>2</sup>, David Madge<sup>2</sup>, Li Xing<sup>2</sup>, Tao Guo<sup>2</sup> and Alán Aspuru-Guzik<sup>3,4,5,6</sup>

- Fibrosis(섬유증)에 관여하는 receptor tyrosine kinase를 저해할 수 있는 저분자 화합물 디자인을 강화학습으로 실행.
- 강화학습으로 만들어낸 신약 후보 물질로 cell-based assay 에서 inhibition 효과가 실제로 있음을 보여줌.(후보들 중 2개)
- 모델을 통해 21일만에 30,000여 개의 저분자 화합물 디자인
- 합성한 사례는 많지만 실제 효과를 본 사례는 매우 적음에 있어 본 논문은 강화학습 신약개발에 대한 가능성을 보여줌

<https://www.nature.com/articles/s41587-019-0224-x>

## 참고 사이트

- <https://kr.mathworks.com/discovery/reinforcement-learning.html>
  - <https://www.comworld.co.kr/news/articleView.html?idxno=49705>
  - <https://davinci-ai.tistory.com/31>
  - <https://media.fastcampus.co.kr/knowledge/data-science/reinforcement/>
  - <http://blog.skby.net/%EA%B0%95%ED%99%94%ED%95%99%EC%8A%B5-reinforcement-learning/>
  - <https://hyeonjiwon.github.io/machine%20learning/ML-1/>
  - <https://koreascience.kr/article/JAKO202214437211866.pdf>
- 
- [http://www.biospectator.com/view/news\\_view.php?varAtcId=6314](http://www.biospectator.com/view/news_view.php?varAtcId=6314)

## 참고 review 논문

<https://www.sciencedirect.com/science/article/pii/S1361841521002395>