



Machine Learning vs. Statistical Model for Prediction Modelling: Application in Medical Imaging Research

예측모형의 머신러닝 방법론과 통계학적 방법론의 비교: 영상의학 연구에서의 적용

Leeha Ryu, MS¹ , Kyunghwa Han, PhD^{2*}

¹Department of Biostatistics and Computing, Yonsei University Graduate School, Seoul, Korea

²Department of Radiology, Research Institute of Radiological Science, Yonsei University College of Medicine, Seoul, Korea

ORCID iDs

Leeha Ryu <https://orcid.org/0000-0002-6575-9531>

Kyunghwa Han <https://orcid.org/0000-0002-5687-7237>

Received August 9, 2022
Revised November 4, 2022
Accepted November 13, 2022

*Corresponding author

Kyunghwa Han, PhD
Department of Radiology,
Research Institute of
Radiological Science,
Yonsei University
College of Medicine,
50-1 Yonsei-ro, Seodaemun-gu,
Seoul 03722, Korea.

Tel 82-2-2228-2394

Fax 82-2-393-3035

E-mail khhan@yuhs.ac

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Clinical prediction models has been increasingly published in radiology research. In particular, as a radiomics research is being actively conducted, the prediction model is developed based on the traditional statistical model, as well as machine learning, to account for the high-dimensional data. In this review, we investigated the statistical and machine learning methods used in clinical prediction model research, and briefly summarized each analytical method for statistical model, machine learning, and statistical learning. Finally, we discussed several considerations for choosing the prediction modeling method.

Index terms Precision Medicine; Medical Imaging; Clinical Decision Rules; Machine Learning

서론

최근 의학 연구에서는 환자 맞춤 의학을 위한 임상 예측 모형 개발과 검증이 활발하게 이루어지고 있다. 진단 또는 예후 예측 목적의 임상 예측 모형에 예측 인자로는 기본적인 환자 특성(나이, 성별 등)과 혈액검사, 뇨검사 등의 결과값과 함께 영상 인자의 역할이 커지고 있다. 따라서, 영상의학 연구 분야에서도 영상 인자를 포함한 임상 예측 모형의 수요가 증가하고 있다(1). 전통적으로 임상 예측 모형은 통계학적 방법론 중 회귀분석에 기반하는 경우가 많다(2). 회귀분석은 한 개 이

상의 독립변수가 종속변수에 미치는 영향을 분석할 수 있는 통계 방법으로, 다양한 종속변수의 형태에 따라 선형, 로지스틱, 생존 회귀분석 등을 활용할 수 있고, 회귀계수를 통해 독립변수가 종속변수에 미치는 영향을 평가할 수 있어 임상 연구에서 유용하게 쓰이는 방법이다.

한편, 최근 영상의학 분야에서 많이 출판되는 라디오믹스나 딥러닝 등 인공지능을 활용한 연구에서도 결국 환자의 질환을 분류하거나 생존, 재발 등의 임상 예후를 예측하는 데에 목적이 있어, 영상의 특성에 대한 분석 및 진단 정확도 평가뿐만 아니라 영상 소견의 환자 예후에 대한 예측력을 보는 연구가 점차 증가하고 있다. 라디오믹스 연구는 주어진 영상에서 전통적인 정성적 특성이 아닌 정량적 특성을 정의하여 질환 분류나 예후를 예측하고자 하는 것으로, 보통 한 개의 영상에서 수십 내지는 수백 개의 특성을 정의하기 때문에 환자 수보다 특성 즉 변수의 개수가 더 많게 된다(3). 이러한 자료에 로지스틱 회귀분석과 같은 기존의 전통적인 통계 방법을 적용하면 독립변수인 라디오믹스 특성과 종속변수 간의 관계에 대한 추정에 실패할 가능성이 높다. 따라서, 고차원 자료를 효율적으로 처리할 수 있는 머신러닝에 기반한 분석을 활용하는 편이다. 유의한 라디오믹스 특성을 먼저 선정하여 고차원이 아닌 형태의 자료로 만든 후 전통적인 통계 문제로 돌아와서 회귀분석에 기반한 분석을 하거나 다수의 영상 특성을 그대로 반영하여 머신러닝에 기반한 분석을 할 수도 있어, 두 가지 접근 중 어떤 것을 하는게 좋을지 고민하게 된다.

본 종설에서는 영상의학 분야의 예측 모형 연구에 사용된 통계학적 방법과 머신러닝 방법들을 조사하여 정리하고, 각 방법론에 대한 설명과 장단점을 살펴보고자 한다. 마지막으로 예측 모형 연구에서 분석 방법 선택에서의 고려사항을 정리해 보고자 한다.

통계학과 머신러닝

기존의 통계학적 모형과 머신러닝(기계학습), 그리고 통계학습(statistical learning)에 대해 비교하는 논문들은 계속 있어 왔지만(4) 분명하게 구분하기는 쉽지 않다. 통계학이란 관심 대상에 대해 관련된 자료를 수집 요약정리하고 이로부터 불확실한 사실에 대한 결론이나 일반적인 규칙성을 추구하는 학문을 일컫는다(5). 따라서 임상 연구에서처럼 표본으로부터 얻은 자료를 이용하여 추론한 연구 결과를 모집단으로 일반화하기 위해서는 추론에 따른 불확실성을 신뢰구간 등으로 제시하고 연구 가설에 대한 지지 정도를 통계학적 분포와 확률에 근거하여 추정하는 과정이 필요하다. 즉 전통적으로 통계학적 방법은 t-분포나 F-분포처럼 알려진 통계학적 이론이나 분석자가 미리 설정한 회귀모형 등의 통계학적 모형에 자료를 적합시켜 가설 검정 또는 연관성 추정을 한다. 한편, 머신러닝은 자료로부터 결과를 추정하는 것이 아니고 자료로부터 자동화된 학습을 통해 모형, 즉 알고리즘을 얻고자 하는 것으로 변수들 간의 연관성 추정보다는 정확한 결과 예측을 목표로 하여 알고리즘을 구축한다(6). 특히 비선형 관계나 고차원 자료에서의 변수 간 교호작용들을 탐색하는 데에 효과적이며, 랜덤 포레스트, 인공신경망 등이 이에 속한다. 통계학습은 벌점화 회귀 모형(penalized regression)에 기반하는 least absolute square shrinkage and selection operator (이하 LASSO), Ridge, Elastic-net, 또는 일반화 가법 모형(generalized additive model) 등이 이에 해당하며(7), 비선형 관계를 탐색할 수 있고 고차원 자료에 효율적이라는 점에서 머신러닝과

유사하나, 과적합(overfitting)을 줄이는 데에 벌점 함수 또는 비선형 함수를 활용할 뿐 통계학적 모형에 기반하여 접근한다는 점에서 머신러닝과는 구분하여 분류한다.

통계학적 모형과 머신러닝을 비교하는 연구들은 다양한 분야에서 출판되어 왔다(4). 각 임상 분야에서 체계적 문헌고찰과 메타분석을 통해 두 방법론의 성능을 비교한 연구들(8-12)과 다양한 변수 선택과 통계 및 머신 러닝 방법들의 적용을 통해 모형을 탐색한 연구들(13, 14)을 참고할 수 있다.

문헌 검색을 통한 영상의학 분야의 예측 모형 연구 사례

영상의학 분야에서 예측모형 구축을 위해 사용된 통계학적 모형과 머신러닝 기법들을 조사하기 위해 관련 논문들을 PubMed와 MEDLINE을 통하여 검색하였다. 문헌 검색은 2018년 1월 1일부터 2022년 7월 8일까지 영상 의학 분야의 학술지에 출판된 영문 논문을 대상으로 수행하였으며, 메타 분석 또는 리뷰 논문은 연구에 포함하지 않았다. 검색 대상 학술지는 영상 의학과 관련성 정도와 2021년 journal impact factor를 종합적으로 고려하여 Journal Citation Reports의 RADIOLOGY, NUCLEAR MEDICINE & MEDICAL IMAGING 분야에서 선정하였다. 최종적으로 *Radiology*, *JACC: Cardiovascular Imaging*, *Medical Image Analysis*, *IEEE Transactions on Medical Imaging*, *Investigative Radiology*, *European Heart Journal-Cardiovascular Imaging*, *Korean Journal of Radiology*, *European Radiology*, 그리고 *AJR. American Journal of Roentgenology* 총 9개의 학술지로 제한하여 논문 검색을 수행하였다. 검색어는 Supplementary Material (in the online-only Data Supplement)과 같다.

문헌 검색 결과를 Table 1에 요약하였다. 사용한 통계모형 또는 머신러닝 방법과 그 수를 요약하였다. 모든 방법 중 로지스틱 회귀분석이 제일 많이 사용되었고(43.8%) 콕스 비례 위험 회귀분석(18.4%)이 뒤를 이었다. 그다음으로는 랜덤 포레스트가 많이 사용되었고(10.5%) 이는 머신 러닝 중 가장 많이 사용된 방법이었다. 각 방법에 대한 간단한 소개는 다음과 같다.

선형 회귀(Linear Regression)

선형회귀는 종속 변수가 연속형인 경우 사용되는 전통적인 통계 모형이다. 회귀 계수를 통해 각각의 설명 변수가 종속 변수에 미치는 영향의 크기를 측정하고 영향의 유의성을 판단할 수 있기 때문에 해석이 용이하다. 선형 회귀는 설명 변수와 종속 변수가 선형 관계를 가지고 있다는 가정 하에 모형을 생성하기 때문에 실제 데이터의 종속 변수와 설명 변수의 관계가 선형이 아닌 경우에는 예측 모형의 정확도가 감소한다. 또한 데이터의 수가 설명 변수의 수에 비하여 충분히 많지 않은 경우 회귀 계수 추정에 실패하여 해석이 불가능해질 수 있다.

로지스틱 회귀(Logistic Regression)

종속 변수의 형태가 예/아니오 와 같이 이분형 변수인 경우에는 로지스틱 회귀분석을 통해 예측 모형 생성이 가능하다. 로지스틱 회귀는 종속 변수의 값을 직접 예측하는 선형 회귀와는 달리, 두 범주 중 특정 한 범주에 속할 확률을 모델링 하는 방법이다. 모형식을 통해 얻은 0과 1 사이의 확률

Table 1. Conventional Statistical Models and Machine Learning Methods as a Prediction Model

Methods	Number of Articles (%)
Conventional statistical models	
Linear regression	92 (8.6)
Logistic regression	468 (43.8)
Logistic regression with LASSO	43 (4.0)
Cox's proportional hazard model	197 (18.4)
Cox's proportional hazard model with LASSO	5 (0.5)
Machine learning methods	
Decision tree	28 (2.6)
Random forests	112 (10.5)
Random survival forests	4 (0.4)
Naïve bayes	11 (1.0)
K-nearest neighbors	16 (1.5)
Support vector machine	84 (7.9)
Linear/quadratic discriminant analysis	8 (0.7)

LASSO = least absolute square shrinkage and selection operator

값을 특정 기준에 따라 두 군으로 나누어 분류 예측을 할 수 있으며, 일반적으로는 0.5를 기준으로 두 군을 분류한다.

콕스 비례 위험 회귀(Cox's Proportional Hazard Regression)

콕스 비례 위험 회귀 모형은 사망이나 재발과 같이 관심 있는 사건의 발생 여부뿐 아니라 발생까지의 시간도 고려하는 경우에서 사건 발생 위험률(hazard rate)을 종속변수로 설정하여 회귀분석을 하는 전통적인 통계 방법이다. 콕스 비례 위험 모형은 준모수적(semi-parametric) 방법으로 생존 자료의 분포를 가정하지 않고도 사건 발생 위험을 추정할 수 있다는 장점이 있다. 두 집단의 사건 발생 위험률의 비(hazard ratio)가 시간에 따라 변하지 않고 일정하다는 비례 위험 가정에 사용 가능하나, 실제 자료에서는 비례 위험 가정을 만족하지 못하는 경우도 있어 확인이 필요하며, 비례 위험 가정을 하지 못하는 경우 다른 통계모형이나 머신러닝 기법 등을 사용할 필요가 있다.

LASSO

일반적으로 여러 개의 설명 변수를 이용하여 종속 변수를 예측하고자 할 때는 다중회귀분석(multivariable analysis) 방법을 사용한다. 하지만 분석 대상자 수에 비해 고려하고자 하는 변수의 개수가 많으면 일반적으로 사용하는 회귀모형에서 회귀 계수에 대한 추정에 실패하고 모형의 예측 성능이 감소할 가능성이 커진다. LASSO는 회귀 계수 추정 시에 회귀계수의 절대값들의 합이 특정 값 이하가 되는 조건 하에 회귀 계수를 추정한다(15). 이렇게 설정함으로써 중요하지 않은 설명변수의 회귀 계수를 0으로 추정하게 하는데 이를 회귀 계수를 축소 추정(shrinkage estimation)한다, 또는 정칙화(regularization) 한 방법이라고도 설명할 수 있다. 회귀 계수가 0이라는 것은 해당 변수의 영향이 없다는 것으로 해당 변수를 모형에서 제외한다는 의미가 되고, 따라서 변수 선택

이 가능해지며 모형의 분산이 감소하여 성능이 좋아지게 된다. 변수 선택 정도를 설정하는 조율 모수(tuning parameter) λ 는 그 값을 0으로 설정하는 경우에는 일반 회귀 모형과 같고, λ 값이 커질수록 축소 효과가 커져서 더 많은 회귀 계수를 0으로 추정하게 된다. LASSO 방법은 여러 모형들과 결합하여 사용이 가능하며, 특히 로지스틱 회귀 모형 또는 콕스 회귀 모형에 자주 사용된다.

의사결정나무(Decision Tree)

의사결정나무는 특정 분류 기준에 따라서 반복적으로 데이터를 나누어 가는 지도학습 방법이다. 전체 데이터가 속한 뿌리 마디(root node)부터 분류 기준에 따라 하위 마디(child node)를 생성하며 데이터를 작은 집단으로 분류해 나가는 방법으로, 그 모습이 나무와 같아 의사결정나무로 불린다. 나무의 각 마디(node)에서는 모든 변수를 탐색하여 데이터를 가장 잘 분할하는 변수를 선택하여 나무의 가장 마지막 지점인 끝마디(terminal node)에 도달할 때까지 나무를 성장시킨다. 각각의 끝마디에 한 개의 데이터만이 속하도록 나무를 최대로 성장시키는 경우에는 과적합이 발생하여 훈련 데이터(training data)는 잘 예측할 수 있으나, 평가 데이터(test data)에서는 좋은 예측 성능을 보이지 못한다. 이러한 과적합 문제를 피하기 위해 가지치기(pruning)를 수행할 수 있다. 의사결정나무는 분류 문제와 회귀 문제 모두에 적용이 가능하며, 도식화가 가능하고 해석이 용이하다는 장점이 있다. 또한 종속 변수와 설명 변수 간의 관계가 비선형으로 복잡한 경우에 적용이 가능하다. 다만 의사결정나무는 예측 정확도가 다른 회귀 및 분류 방법에 미치지 못한다는 단점이 있다.

랜덤 포레스트(Random Forests)

랜덤 포레스트는 의사결정나무의 예측 정확도 향상을 위해 고안된 방법으로, 여러 결과를 종합하여 최종 결정을 하는 앙상블 방법 기반의 기계학습 방법이다. 붓스트랩(bootstrap)을 이용하여 여러 번의 재표본추출로 생성한 여러 훈련 데이터로 다수의 의사결정나무를 생성한 후 각 나무에서 얻어진 결과를 종합하여 최종 결과를 도출한다. 하나의 나무에서 각 마디를 나눌 때는 무작위로 p 개의 전체 변수 중 m ($m \leq p$) 개의 후보 변수를 선정하여 최종적으로 마디를 나눌 변수를 선택한다. 즉 랜덤 포레스트는 분석 대상자와 변수 선정 과정에 임의성(randomness)을 부여함으로써 각 나무 간의 연관성을 감소시키며, 연관성이 적은 다수의 나무에서 얻어진 결과를 종합하여 최종 결과를 도출하기 때문에 과적합을 피할 수 있다는 장점이 있다. 최종 결과 예측을 위해서는 분류 문제의 경우 다수결의 원칙을 사용하여 전체 나무에서 가장 많이 분류된 범주로 예측하며, 회귀 문제의 경우에는 각 나무에서 얻어진 결과를 평균 내어 예측한다. 랜덤 포레스트는 상대적으로 대상자 수도 많으면서 전체 변수의 개수가 많은 경우에 효과적이며(16), 가장 대표적으로 사용되는 나무 기반 방법이다.

랜덤 서바이벌 포레스트(Random Survival Forests)

랜덤 포레스트를 생존 자료에 적용하기 위해서 랜덤 서바이벌 포레스트를 사용할 수 있다. 랜덤 포레스트에 기반한 방법으로 두 하위 마디의 생존 시간의 차이가 최대가 되도록 상위 마디를 나눈

다. 생존 자료에서 주로 사용되는 콕스 비례 위험 모형과는 달리 비례 위험 가정을 만족하지 않아도 사용 가능하고 비선형 관계나 변수 간 교호작용을 반영할 수 있다는 장점이 있다. 영상의학 분야에서는 생존에 대한 예후 예측을 목적으로 하는 경우 중 Radiomics와 같이 변수가 상대적으로 많은 자료이거나 콕스 비례 위험 가정을 하기 어려운 자료에서 예측 모형으로 사용할 수 있다(17).

나이브 베이즈(Naïve Bayes)

나이브 베이즈는 사전 확률을 이용하여 사후 확률을 추론하는 베이즈 정리에 기반한 데이터 분류기이다. 데이터가 분류될 범주가 사전에 정해진 경우에 사용 가능하며 각각의 범주에 대해서 설명 변수들이 독립이라고 가정한다. 실제 데이터의 경우 이러한 조건부 독립을 만족하지 않는 경우가 많지만, 그럼에도 나이브 베이즈는 좋은 성능을 보인다. 설명변수들의 조건부 독립이라는 가정으로 인해 나이브 베이즈는 분석 대상자 수가 설명 변수의 개수에 비해 많지 않은 경우에도 잘 작동한다(18, 19).

K 최근접 이웃 방법(K-Nearest Neighbors Method)

K 최근접 이웃 방법은 한 관측치에서 가장 가까운 K개의 자료가 가장 많이 속한 범주로 관측치를 분류하는 방법으로, 모형 성능을 탐색하며 최적의 K를 선정할 수 있다. 구축용 데이터에서 가장 가까운 데이터만 탐색하는 경우, 즉 K가 1인 경우 바로 옆 데이터에게만 영향을 받기 때문에 각 범주를 분류하는 경계가 매우 유연하나, K가 커질수록 여러 개의 주변 데이터를 고려하여 분류하기 때문에 직선에 가까운 형태를 가지게 된다. 실제 데이터에서는 데이터가 가지는 분포의 형태를 알 수 없는 경우가 대부분이나, K 최근접 이웃은 매우 단순한 알고리즘을 사용하므로 데이터의 분포를 가정할 수 없는 상황에서도 분류가 가능하다(19, 20).

서포트 벡터 머신(Support Vector Machine)

분류 문제에서 주로 사용되는 서포트 벡터 머신은 각 범주를 잘 분류하는 경계를 찾는 기계학습 방법이다. 각 범주에 속하는 자료와 경계의 거리 중 가장 짧은 거리를 마진(margin)이라 하면, 데이터의 범주를 구분하는 무수히 많은 경계들 중 마진이 가장 큰 경계를 선택한다. 즉, 각 범주에서 가장 멀리 떨어진 경계를 찾는 것이고 이 경계를 서포트 벡터라고 부른다. 이는 선형과 비선형 형태 모두 고려할 수 있다(19).

선형/이차 판별 분석(Linear/Quadratic Discriminant Analysis)

관심 있는 종속 변수가 범주형 변수일 때, 범주 간 특성의 차이가 크지 않는 경우 로지스틱 예측 모형의 예측 성능이 감소한다. 하지만 각 범주에 속한 데이터가 잘 분리되어 있지 않는 경우에도 잘 작동하는 판별분석은 각 범주에 속한 데이터가 정규분포 또는 가우스 분포를 따른다는 가정하에 각 범주 내의 분산은 작아지고 범주 간의 분산은 커지도록 데이터를 새로운 축에 사영한 후 분류하는 경계를 찾는 방법이다. 선형 판별분석에서는 각 범주들의 동일한 분산을 갖는다고 가정하고 데이터 분류 경계가 직선 형태를 가지게 되고, 이차 판별분석에서는 각 범주의 분산을 각각 다

르게 가정하여 유연한 형태의 경계를 가진다. 따라서 선형 판별분석은 이차 판별분석 보다 낮은 분산과 더 좋은 예측 성능을 가지나 높은 편향(bias)을 가지게 된다. 구축용 자료의 개수가 적은 경우에는 선형 판별분석을 사용하는 것이 좋으며 구축용 자료가 많거나 각 범주가 동일한 분산을 갖는다는 가정이 타당하지 않는 경우에는 이차 판별분석을 사용하는 것이 더 좋은 예측 성능을 보일 수 있다(19).

예측 모형 연구에서 분석 방법 선택 방안

앞서 살펴보았듯이 예측 모형을 구축하는 데에는 다양한 모형을 적용해 볼 수 있어 선택의 문제에 부딪히게 된다. 결과 변수의 형태에 따라 연속형 수치의 예측, 분류, 또는 생존 등 예측 모형의 목적을 정할 수 있고 여기서 1차적으로 모형을 선정할 수 있다. 그다음으로는, 예측 변수와 결과 변수의 관계의 형태, 예측 변수들 간의 상호 관계 등에 따라 모형을 선정할 수 있다. 선형 모형으로도 충분하다면 임상 연구에서 쉽게 접근 가능한 통계학적 모형을 사용할 수 있지만, 선형 관계로 가정하기 어려운 경우나 복잡한 변수들 간의 상호 관계를 반영하고자 한다면 머신러닝이 효과적일 수 있다. 대상자 수, 특히 분류나 생존 문제에서는 관심 있는 사건이나 그 반대 중 적은 쪽의 대상자 수와 변수의 개수의 비 또한 모형을 선택하는 데에 고려할 수 있다. 대상자 수에 비해 변수의 개수가 훨씬 많은 자료의 형태라면 통계학습(LASSO 등) 방법이 효과적일 수 있고, 대상자 수 또한 많은 편이라면 자료 재추출을 활용하는 머신러닝이 효과적일 수 있다(16).

이와 같이 연구 목적과 자료 형태에 따라 분석방법을 선정한 후에는 모형에 넣을 예측 인자를 정해야 하는데, 통계학적 모형에 대해서는 경험적으로 events per variable (이하 EPV)을 기준으로 하여 EPV가 10 이상이면 모형이 안정적으로 추정된다는 점이 알려져 있어 이를 참조하는 편이다. 예를 들어 관심 있는 범주에 해당하는 대상자 가 30명이라면 $30/10 = 3$ 개의 예측 인자를 넣는 것이 적절하다고 판단하는 것이다. 하지만, 이 역시 특정 상황에서의 모의실험 연구(21, 22)에 기반한 결과이므로 적용할 때 주의해야 하며, 최근에는 이 기준을 좀 더 완화하거나(23) 계획하는 예측 모형의 실제 분포를 설정하여 대상자 수와 예측 인자의 개수를 정확하게 추정하자는 제안도 나오고 있다(24). 또한, 예측 모형은 개별 예측 변수의 효과 추정이나 유의성 검정보다는 여러 개의 예측 변수를 함께 고려했을 때의 정확한 예측을 중요시하므로 EPV를 참조하는 원칙보다는 다각도에서의 모형 성능 평가를 통해 최종 모형을 확정하는 것이 더 중요하다고 할 수 있겠다.

토론 및 결론

본 종설에서는 예측모형을 구축할 때 사용하는 통계학적인 방법과 머신러닝에 대해 간단히 소개하고 영상의학 분야에서 사용된 경우들을 조사 및 정리하였다.

의료영상은 다양한 특성을 내포하고 있고 이를 정량화한 라디오믹스 분야가 활발하게 진행 중이며, 의학 자료 역시 빅데이터 형태의 종합적인 자료를 얻을 수 있어 이러한 방대한 자료를 예측 모형에 효과적으로 반영할 수 있는 방법이 필요하다. 모형의 성능을 평가할 때에는 모형의 예측에

대한 정확도뿐 아니라 정확도에 대한 신뢰 가능성, 모형의 일반화 가능성 등 다양한 관점에서의 접근이 필요하다. 분석 방법의 결정에 있어 부족한 정답은 없지만 분석대상자 수와 변수의 개수를 고려하여 효율적인 방법을 시도해 볼 수 있다. 탐색적인 관점에서 다양한 방법을 모두 적용해 보고 최적의 성능을 보이는 경우를 선정할 수 있으나 그러한 과정을 통해 선정한 모형이 외부 자료나 앞으로 적용할 자료에서도 비슷한 성능을 보일지 등을 고려하여 로버스트(robust)한 모형을 개발하고 검증하는 과정이 필요하다.

본 종설에서는 모형의 구축에서 고려할 점들을 주로 소개하였으나 구축된 모형을 검증하는 문제 또한 매우 중요하다. 외부 검증이 되지 않은 모형은 신뢰도가 떨어져서 활용되기 어렵다. 특히 활용 가능성에 있어 기존의 통계 모형은 지면상에 출판된 모형식(회귀계수, 오즈비 등)만으로도 활용 가능하였으나 머신러닝을 이용한 예측 모형은 해당 알고리즘을 분석 소프트웨어나 웹사이트 등을 통해 공유할 수 있어야 외부 검증 및 활용이 가능하므로 이에 대한 고려가 필요하다. 저널에 예측 모형 연구를 투고하는 경우 해당 보고 가이드라인인 Transparent Reporting of a multi-variable prediction model for Individual Prognosis Or Diagnosis (이하 TRIPOD) statement (2, 25)를 참고하는 것을 권장하며, 군집화된 자료(다기관, 다판독자, 한 환자당 여러 병변이 있는 경우 등)을 고려하는 TRIPOD clustered data (26), 머신러닝 등 인공지능을 활용한 예측 모형에 대한 가이드라인인 TRIPOD-AI statement (27)도 출간 준비 중으로 알려져 있어 추후 유용하게 활용할 수 있을 것으로 기대한다.

Author Contributions

Conceptualization, H.K.; data curation, R.L.; formal analysis, R.L.; methodology, all authors; supervision, H.K.; writing—original draft, all authors; and writing—review & editing, H.K.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Funding

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1I1A1A01059893).

REFERENCES

1. Han K, Song K, Choi BW. How to develop, validate, and compare clinical prediction models involving radiological parameters: study design and statistical methods. *Korean J Radiol* 2016;17:339-350
2. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55-63
3. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016; 278:563-577
4. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods* 2018;15:233-234
5. Rosner B. *Fundamentals of biostatistics*. Boston, MA: Cengage Learning 2015
6. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med* 2001;23:89-109
7. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. New York: Springer 2009
8. Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ* 2021;

375:n2281

9. Yang C, Kors JA, Ioannou S, John LH, Markus AF, Rekkas A, et al. Trends in the conduct and reporting of clinical prediction model development and validation: a systematic review. *J Am Med Inform Assoc* 2022;29:983-989
10. Sun Z, Dong W, Shi H, Ma H, Cheng L, Huang Z. Comparing machine learning models and statistical models for predicting heart failure events: a systematic review and meta-analysis. *Front Cardiovasc Med* 2022;9:812276
11. Sufriyana H, Husnayain A, Chen YL, Kuo CY, Singh O, Yeh TY, et al. Comparison of multivariable logistic regression and other machine learning algorithms for prognostic prediction studies in pregnancy care: systematic review and meta-analysis. *JMIR Med Inform* 2020;8:e16503
12. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol* 2019;110:12-22
13. Leger S, Zwanenburg A, Pilz K, Lohaus F, Linge A, Zöphel K, et al. A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling. *Sci Rep* 2017;7:13206
14. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine learning methods for quantitative radiomic biomarkers. *Sci Rep* 2015;5:13087
15. Pavlou M, Ambler G, Seaman S, De Iorio M, Omar RZ. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Stat Med* 2016;35:1159-1177
16. Park JE, Park SY, Kim HJ, Kim HS. Reproducibility and generalizability in radiomics modeling: possible strategies in radiologic and statistical perspectives. *Korean J Radiol* 2019;20:1124-1137
17. Bae S, Choi YS, Ahn SS, Chang JH, Kang SG, Kim EH, et al. Radiomic MRI phenotyping of glioblastoma: improving survival prediction. *Radiology* 2018;289:797-806
18. Tan CO, Lam S, Kuppens D, Bergmans RHJ, Parameswaran BK, Forghani R, et al. Spot and diffuse signs: quantitative markers of intracranial hematoma expansion at dual-energy CT. *Radiology* 2018;290:179-186
19. Eun NL, Kang D, Son EJ, Park JS, Youk JH, Kim JA, et al. Texture analysis with 3.0-T MRI for association of response to neoadjuvant chemotherapy in breast cancer. *Radiology* 2020;294:31-41
20. Johnson KM, Johnson HE, Zhao Y, Dowe DA, Staib LH. Scoring of coronary artery disease characteristics on coronary CT angiograms by using machine learning. *Radiology* 2019;292:354-362
21. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373-1379
22. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol* 1995;48:1503-1510
23. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol* 2007;165:710-718
24. Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368:m441
25. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1-W73
26. TRIPOD. TRIPOD clustered data. Available at: <https://www.tripod-statement.org/clustered/>. Accessed July 8, 2022
27. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open* 2021;11:e048008

예측모형의 머신러닝 방법론과 통계학적 방법론의 비교: 영상의학 연구에서의 적용

유리하¹ · 한경화^{2*}

최근 영상의학 연구 분야에서 영상 인자를 포함한 임상 예측 모형의 수요가 증가하고 있고, 특히 라디오믹스 연구가 활발하게 이루어지면서 기존의 전통적인 회귀 모형뿐만 아니라 머신러닝을 사용하는 연구들이 많아지고 있다. 본 종설에서는 영상의학 분야에서 예측 모형 연구에 사용된 통계학적 방법과 머신 러닝 방법들을 조사하여 정리하고, 각 방법론에 대한 설명과 장단점을 살펴보고자 한다. 마지막으로 예측 모형 연구에서 분석 방법 선택에서의 고려사항을 정리해 보고자 한다.

¹연세대학교 일반대학원 의학전산통계학협동과정,

²연세대학교 의과대학 영상의학교실, 방사선외과학연구소, 의료영상데이터사이언스센터