



Application of Machine Learning in Rhinology: A State of the Art Review

Myeong Sang Yu

Department of Otorhinolaryngology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

비과 영역에서 머신러닝의 적용

유명상

울산대학교 의과대학 서울아산병원 이비인후과학교실

Received July 2, 2020

Accepted July 30, 2020

Address for correspondence

Myeong Sang Yu, MD, PhD
Department of Otolaryngology,
Asan Medical Center,
University of Ulsan
College of Medicine,
88 Olympic-ro 43-gil, Songpa-gu,
Seoul 05505, Korea
Tel +82-2-3010-3710
Fax +82-2-489-2773
E-mail dryums@gmail.com

The revolutionary development of artificial intelligence (AI) such as machine learning and deep learning have been one of the most important technology in many parts of industry, and also enhance huge changes in health care. The big data obtained from electrical medical records and digitalized images accelerated the application of AI technologies in medical fields. Machine learning techniques can deal with the complexity of big data which is difficult to apply traditional statistics. Recently, the deep learning techniques including convolutional neural network have been considered as a promising machine learning technique in medical imaging applications. In the era of precision medicine, otolaryngologists need to understand the potentialities, pitfalls and limitations of AI technology, and try to find opportunities to collaborate with data scientists. This article briefly introduce the basic concepts of machine learning and its techniques, and reviewed the current works on machine learning applications in the field of otolaryngology and rhinology. Korean J Otorhinolaryngol-Head Neck Surg 2020;63(8):341-9

Key Words Artificial intelligence · Deep learning · Machine learning · Medicine · Otolaryngology · Rhinology.

서 론

인공지능(artificial intelligence)은 과학 기술이 발전하면서 그 개념도 함께 변하고 있기 때문에 정의를 내리기 쉽지 않지만, 간단히 설명하면 인간의 지능을 모방하여 사람이 하는 것과 같이 복잡한 일을 할 수 있는 기계를 만드는 것을 말한다.¹⁾ 컴퓨터 하드웨어와 공학 등 관련 분야가 빠르게 발전하고 엄청난 양의 데이터를 빠른 시간 내에 효과적으로 다룰 수 있게 되면서 인공지능 기술은 이제 소셜 속에 등장하는 환상이 아니라 현실이 되어가고 있다. 최근에는 여러 인공지능 기술 중에서도 인간이 만든 프로그램 없이 컴퓨터가 스스로

데이터를 학습하여 그 의미를 해석해 내는 머신러닝(machine learning)과 딥러닝(deep learning) 기술이 실제 문제 해결에 적용되기 시작하면서 인공지능 성능의 비약적인 발전을 가져왔다. 의료 분야에서는 전자의무기록(electrical medical record)이나 의료 영상 전송 시스템(picture archiving and communication system) 등을 통해 기존에는 볼 수 없었던 빅데이터를 축적할 수 있게 되고, 이러한 빅데이터를 이용한 인공지능 분야의 기술이 광범위하게 도입되어 일부는 임상에서 적용되기 시작했다. 이러한 흐름은 향후 의료계 전체에도 큰 변화를 가져올 것으로 보인다.

본 종설에서는 머신러닝과 딥러닝의 기본 개념들과 대표적인 알고리즘 모델들을 간략히 설명하고, 최근 이비인후과와 비과 영역에서 머신러닝을 적용한 연구들을 소개하고자 한다.

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.

머신러닝(Machine Learning)

인공지능은 포괄적으로 ‘인간의 지능을 인공적으로 모사하는 것’을 의미하며, 여러 인공지능 방법 중 크게 ‘규칙 기반’으로 사전에 규칙을 정해서 판단, 추론하는 방법과 데이터를 통해서 모형을 학습하게 하는 ‘머신러닝(기계학습)’ 분야로 나눌 수 있다. 규칙을 기반으로 하는 방법은 인간이 미리 정해둔 규칙을 통해서 판단을 하게 되지만, 머신러닝은 데이터를 학습 알고리즘에 집어넣어서 그러한 규칙들을 스스로 찾을 수 있게 만드는 알고리즘 방법이다. 딥러닝은 이러한 머신러닝의 여러 방법론 중의 하나로 인공 신경망(artificial neural network, ANN)의 한 종류라고 할 수 있다(Fig. 1).

머신러닝은 1959년에 처음 소개될 정도로 오래된 개념이지만, 최근에 들어서야 여러 분야에서 널리 이용되고 있는데, 이는 고속 인터넷의 보편화 및 컴퓨터 하드웨어의 발전, 그리고 무엇보다 빅데이터의 등장이 가장 큰 이유라고 할 수 있다. 머신러닝의 학습에는 기본적으로 많은 양의 데이터가 필요하고, 데이터의 양에 따라 머신러닝의 성능이 좌우되므로 충분한 데이터의 확보는 인공지능 개발에서 가장 기본적인 요소라고 할 수 있다. 데이터의 질(quality) 역시 머신러닝에서 중요한 요소이다. 학습에 필요한 조건을 갖추지 못한 오류가 있는 데이터는 머신러닝의 성능을 저하시키고 결과에 영향을 끼칠 수 있으므로, 학습 이전에 반드시 데이터의 질을 평가하는 과정이 수반되어야 한다.

머신러닝과 기존 통계학의 차이

머신러닝과 기존의 전통적인 통계학(traditional statistics)은 데이터를 통해서 문제를 해결한다는 면에서 서로 비슷하다. 실제로 선형 회귀(linear regression)와 같이 머신러닝의 많은 기법들이 통계학의 기법들과 상당히 유사하다. 다만 목표와 전략에서 서로 차이가 있다고 할 수 있다. 전통적 통계학은 일반적으로 작은 데이터셋(data set)을 다루며, 결론에

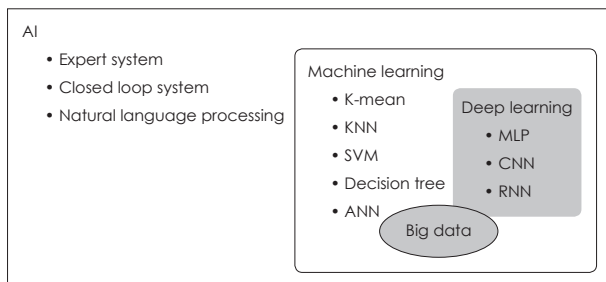


Fig. 1. Relationship between AI, machine learning, and deep learning. AI: artificial intelligence, KNN: k-nearest neighbors, SVM: support vector machine, ANN: artificial neural network, MLP: multiple layer perceptron, CNN: convolutional neural network, RNN: recurrent neural network.

이르기까지 여러 인자들 간의 상관관계에 대한 추론(inference)과 가정(assumption)을 중시한다. 하지만 머신러닝은 빅데이터에 보다 적합하며, 기존의 데이터를 바탕으로 알 수 없는 부분에 대한 정확한 예측(prediction)에 집중한다.²⁾ 두 정부 암의 생존율을 예로 들면, 기존의 통계학에서는 생존율에 영향을 미치는 여러 인자들과의 상관관계를 분석하는 데 초점이 맞춰져 있다면, 머신러닝에서는 기존의 데이터를 통해서 새로운 환자의 생존율을 정확하게 예측하는 데 집중한다고 할 수 있다.

머신러닝을 이용한 예측 모델링의 진행과정

머신러닝을 이용한 예측 모델링(predictive modeling)은 일반적으로 다음과 같은 과정을 거치게 된다.

- 1) 데이터 수집(data collection)
- 2) 데이터 전처리(data preprocessing): 데이터를 정제해서 머신러닝 모델의 입력에 적합한 형태로 바꿔주거나, 데이터의 기존 속성(feature)을 조합해서 새로운 속성을 생성
- 3) 탐색적 데이터 분석(exploratory data analysis): 데이터의 특징을 찾고, 숨겨진 패턴을 발견
- 4) 모델 선택(model selection): 주어진 문제와 데이터에 맞는 적절한 모델과 속성을 선택
- 5) 평가 및 적용(evaluation & application): 만들어진 머신러닝 모델의 성능을 평가하고, 모델을 활용하여 새로운 데이터에 대한 예측을 실행

고전적인 머신러닝 알고리즘의 종류

머신러닝은 지도학습, 비지도학습과 강화학습의 세 가지 종류로 나눌 수 있다. 지도학습(supervised learning)은 입력값(input)과 함께 정답값(label)을 주고 학습을 시키는 방법으로, 알고리즘을 실행하면서 얻어진 결과와 정답값을 함께 비교하면서 반복적으로 훈련해 나가는 방법이다. 현재 이용되고 있는 다수의 머신러닝 알고리즘이 여기에 해당하며, 분류(classification)와 회귀(regression) 등의 문제를 해결하는 데 이용된다. 이에 반해 비지도학습(unsupervised learning)은 정답값이 없이 입력값만을 이용하여 학습을 시키게 되며, 하위 범주로는 클러스터링(clustering), 차원 축소(dimensionality reduction) 등이 있다. 마지막으로 강화학습(reinforcement learning)은 결과값 대신 어떤 작업을 잘 수행했을 때 보상(reward)을 주는 방법으로 학습을 시킨다(Fig. 2).

딥러닝(Deep Learning) 소개

딥러닝은 머신러닝의 여러 방법론 중의 하나로 인공 신경망

(ANN)이 발전된 개념이라고 할 수 있다. 1943년에 처음 소개된 인공 신경망은 두뇌의 신경세포, 즉 뉴런이 연결된 형태를 모방한 모델로, 인간의 신경 구조를 복잡한 스위치들이 연결된 네트워크로 표현할 수 있다는 개념에서 유래되었다.³⁾ 그리고 1958년 Rosenblatt⁴⁾은 퍼셉트론(perceptron)이라는 선형 분류기 모델을 제안했는데, 이는 입력(input)과 가중치(weight)들의 곱을 모두 더한 뒤 활성화 함수(activation function)를 적용해서 그 값이 0보다 크면 1, 0보다 작으면 -1을 출력하는 구조였다(Fig 3A). 현재 사용하는 딥러닝 모델도 근본적으로 퍼셉트론과 같은 구조라고 할 수 있다. 다만 이 구조를 여러 개의 노드(node)와 다층의 레이어(layer)로 확장했다는 점에

서 차이가 있을 뿐이다. 하지만 단순한 퍼셉트론 만으로는 복잡한 문제를 풀 수 없었기 때문에 은닉계층(hidden layer)이라는 중간층을 추가하여 한계를 극복한 다층 퍼셉트론(multi-layer perceptron)이 소개되었다(Fig 3B). 여기에서 퍼셉트론을 여러 층으로 쌓을수록 학습하기 어려운 문제가 제기되었는데, 순방향(feed forward) 연산 후 예측 값과 정답 사이의 오차를 후방(backward)으로 다시 보내주면서 학습시키는 방법인 역전파 알고리즘(backpropagation algorithm)으로 해결되었다.⁵⁾ 하지만 복잡한 문제의 해결을 위해 깊은 층수를 쌓을 경우, 역전파 학습과정에서 데이터가 사라져 학습이 잘 되지 않는 현상이 발생하면서 인공 신경망은 정체기를 겪었다.

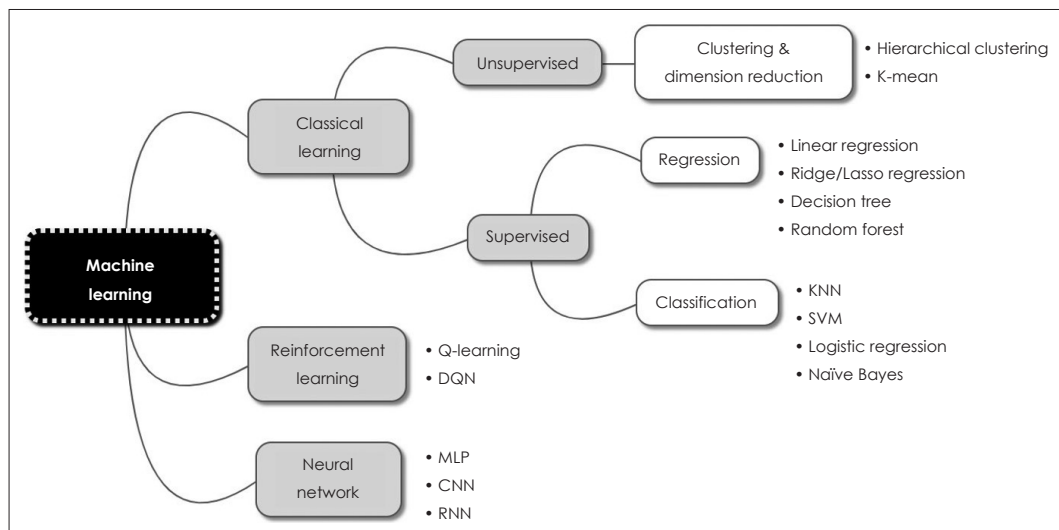


Fig. 2. Commonly used machine learning algorithms. KNN: k-nearest neighbors, SVM: support vector machine, MLP: multiple layer perceptron, CNN: convolutional neural network, RNN: recurrent neural network, DQN: deep Q network.

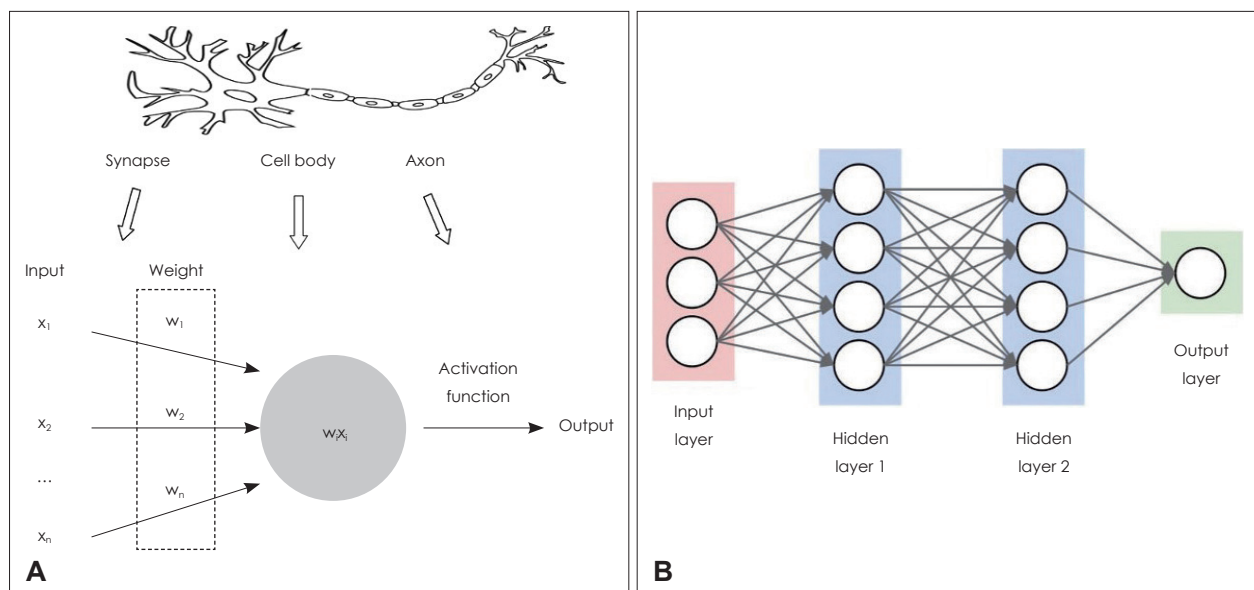


Fig. 3. A conceptual picture of a simple perceptron (A) and artificial neural network (B).

2000년대에 들어서면서 Nair 등⁶⁾에 의해 사전 학습(pre-training)이나 ReLU와 같은 새로운 활성화함수의 적용, 그리고 학습 도중에 고의로 데이터를 누락시키는 방법(dropout)⁷⁾ 등으로 이러한 문제들이 해결되면서 깊은 층을 갖는 신경망 학습(deep neural network)이 가능해졌고, 이후에 인공 신경망은 ‘딥러닝(deep learning)’이라는 새로운 이름으로 불리게 되었다. 딥러닝이 최근 널리 유행하게 된 이유에는 기존 인공 신경망의 한계를 극복한 알고리즘의 개발과 신경망 학습에 필요한 막대한 학습데이터의 축적, 그리고 신경망 학습에 적합한 컴퓨터 하드웨어의 발전 등을 들 수 있다.

딥러닝과 머신러닝 비교

딥러닝은 스스로 데이터의 특성을 찾아내어 스스로 학습을 한다는 점에서 기존의 머신러닝과 가장 큰 차이가 있다. 기존의 머신러닝 기법에서는 사람이 직접 추출하고 분석한 데이터의 특성을 기반으로 학습하여 문제를 해결하였다. 학습 특성은 수행하는 작업에 따라 종류가 많고 어떠한 특성들을 선택하는가에 따라 모델 성능의 한계가 있었다. 하지만 딥러닝을 이용한 학습에서는 스스로 이러한 특성들을 찾아내고 특성 학습을 수행한다. 이밖에도 딥러닝과 기존의 머신러닝 기법 사이에는 여러 차이점들이 있다(Table 1).

대표적인 딥러닝 알고리즘: CNN과 RNN

최근 널리 이용되는 딥러닝 모델에는 합성곱신경망(convolutional neural network, CNN)과 순환 신경망(recurrent neural network, RNN)이 있다. 여기에서는 두 모델의 기본개

념만을 간략히 소개하겠다.

CNN은 이미지를 인식하여 분류하는 데 주로 사용되는 모델로 의료 분야의 영상 분석에 이용되는 대표적인 딥러닝 기법이라고 할 수 있다. 딥러닝은 그 특성상 대상 영상의 특성과 상관없이 일반적인 모델링이 가능하다. 특히 신경망의 계층적 구조는 영상의 일반적인 특성들뿐만 아니라 영상의 색이나 테두리 선과 같은 하위 수준의 특성들까지 스스로 추출하여 학습할 수 있다. CNN에서는 이미지의 위치, 크기, 각도 변화 등에 의해 왜곡되는 것과 상관없이 이미지를 인식하기 위해 컨볼루션(convolution)과 풀링(pooling) 과정을 반복적으로 수행하면서 이미지 데이터에서 추상화된 정보를 추출한다. CNN은 크게 두 부분으로 구성되는데, 초반에는 컨볼루션과 풀링 과정이 반복적으로 존재하고, 마지막 부분에는 다층 퍼셉트론 구조의 fully connected layer를 붙여서 분류기(classifier)의 형태를 갖게 된다. 컨볼루션 단계에서는 입력된 이미지에 여러 가지 필터(filter)를 사용하여 이미지의 특징을 도출하고, 풀링 단계에서는 이미지의 특징은 유지하면서 이미지의 크기를 줄이는 역할을 한다. 이런 과정을 여러 차례 반복해서 하나의 이미지에서 그 이미지의 개별 특징들을 보여주는 다양한 작은 이미지들을 만들어 내게 된다(Fig. 4). 그리고 Inception이나 ResNet과 같이 기존의 잘 학습된 CNN 모델의 내부에 있는 필터셋(filter set)들은 다른 문제 해결에도 사용될 수 있는데 이를 전이 학습(transfer learning)이라고 한다. 이를 잘 활용하면 적은 수의 데이터로도 우수한 성능의 CNN 모델을 만들 수 있으며, 최근 의료 분야에서 발표되고 있는 딥러닝을 이용한 이미지 분석 연구들은 대부분 이

Table 1. The differences between deep learning and machine learning

	Deep learning	Machine learning
Feature extraction	Computer self-extraction and analysis	Human intervention
Data dependency	A large amount of data	Relatively small data
Hardware dependency	High-end machines	Low-end machines
Problem solving approach	End to end	Staged
Execution time	A long time to train	Much less time to train
Interpretability	Difficult to interpret (black box)	Easy to interpret

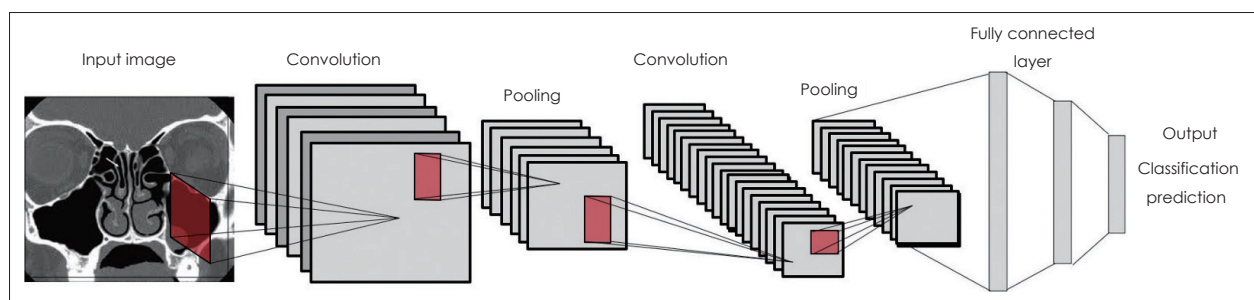


Fig. 4. A conceptual explanation of CNN. CNN consists of a number of convolutional and pooling layers, and then followed by one or more fully connected layers as in a standard multilayer neural network. CNN: convolutional neural network.

러한 CNN 기법과 전이 학습을 이용하고 있다.⁸⁻¹⁰⁾

RNN은 시계열 데이터와 같이 시간의 흐름에 따라 변화하는 데이터를 학습하기 위한 인공 신경망으로 내부에 자기 자신으로 되돌아 오는 층(layer)을 가지고 있다. 따라서 과거의 출력 데이터를 재귀적으로 참조한다. 여기에서 ‘재귀’는 현재의 결과와 이전 결과와 연관성을 가진다는 의미이다. RNN은 음성인식, 기계번역, 이미지 설명 등 여러 분야에서 활용되는데, 대표적인 예가 단어나 문장을 완성해 주는 알고리즘이다. 이전에는 여러 기술적인 문제로 깊은 RNN은 학습이 불가능했으나, 최근 장단기 기억(long short-term memory, LSTM)¹¹⁾ 모델과 gated recurrent unit¹¹⁾ 모델이 발표되면서 가능하게 되었다. 최근 의료 분야에서도 시계열을 이용하는 대부분의 딥러닝 연구들은 RNN의 한 종류인 LSTM 모델을 이용하고 있다.¹²⁾

딥러닝의 한계

딥러닝 모델들이 기존의 머신러닝이나 인공 신경망의 한계를 뛰어 넘으면서 의료 분야를 포함한 많은 분야에서 활용이 크게 증가하고 있지만 아직 해결이 필요한 제한점들이 있기 때문에 기존의 머신러닝 모델들을 완전히 대체하지는 못하고 있다. 딥러닝이 가지고 있는 한계는 첫째, 딥러닝 모델에서는 결과가 잘 나오더라도 모델이 어떻게 그 결과를 도출하였는지에 대한 해석을 할 수는 없다. 그래서 이를 블랙박스(black box) 모델이라고 부르기도 한다. 예를 들면, 환자의 데이터를 기반으로 딥러닝 모델이 이 환자에게 필요한 처방을 예측했다 하더라도, 왜 그런 처방이 나왔는지 설명하지 못하기 때문에 의사와 환자 모두 그 결과를 받아들이기 힘들 수 있다. 둘째, 딥러닝 모델을 학습시키기 위해서는 일반적으로 많은 양의 데이터가 필요하기 때문에, 문제에 따라 다르지만 데이터가 적은 경우에는 딥러닝 모델을 적용하기 어려운 경우가 많다. 특히 임상에서는 여러 제한들 때문에 딥러닝 모델을 만들기 위해 필요한 데이터의 양을 충족시키지 못할 때가 많다. 셋째, 딥러닝 모델을 학습시키기 위해서는 엄청난 계산량이 필요하며, 이를 감당하기 위해서 고사양의 컴퓨터 하드웨어와 상대적으로 긴 학습 시간을 요구한다. 따라서, 결과의 해석이 필요하거나, 데이터의 양이 적은 경우, 그리고 짧은 학습 시간이 필요할 때는 기존의 머신러닝 모델이 딥러닝 모델보다 효과적일 수 있다.

임상에서 머신러닝의 적용

머신러닝이나 딥러닝과 같은 인공지능 기술은 의료 분야의 여러 분야에서 활용될 수 있다. 특히 최근 딥러닝 기법의 발달로 이미지 분석의 성능이 높아지면서 영상 분야에서 많이

적용되어 많은 연구가 진행 중이며 일부는 실제 임상에 사용되고 있다. 영상 분야에서 머신러닝의 역할은 크게 이미지를 통해 특정 질환 여부를 구분할 수 있는 분류, 이미지에서 특정 장기나 병변을 자동으로 분리할 수 있는 분할(segmentation)과 컴퓨터보조검출(computer-assisted detection) 방식의 병변 자동 검출, 그리고 저품질의 영상에서 고품질의 영상과 유사한 영상을 만들어내는 모의(simulation) 등이 있다.¹³⁾ 개인별 맞춤 치료를 추구하는 정밀의료(precision medicine)가 의료 분야의 화두가 되고 있는 현 상황에서, 빅데이터의 분석을 통해 치료의 결과 등을 미리 예측할 수 있는 분류 분야에서 가장 활발하게 머신러닝을 이용한 연구가 진행 중이다.

이비인후과 영역에서 머신러닝을 이용한 연구

이비인후과 분야도 다른 의료 분야와 마찬가지로 머신러닝이나 딥러닝을 이용한 연구들이 진행 중이다. 그 중에서도 두경부 영역에서 머신러닝을 이용한 연구가 가장 활발한데, CNN 모델을 이용하여 내시경이나 CT, MRI 같은 영상에서 암 병변 부위를 자동으로 감지하여 전체 종양의 부피를 계산하거나,^{14,15)} 후두내시경에서 병변에 따라 자동으로 질환을 분류하기도 한다.¹⁶⁻¹⁹⁾ 이와 같이 이미지나 영상만을 가지고 질환을 자동 진단(automated detection)하기 위한 연구들이 광학생검(optical biopsy)이라는 이름으로 시행 중이다.²⁰⁾ 이밖에 머신러닝 알고리즘을 이용하여 두경부암 환자의 생존율이나 합병증 발생, 경부 림프선 전이, 재발 여부 등을 예측하는 연구들도 시행되고 있다.²¹⁻²⁸⁾ 이와 영역에서는 머신러닝 기법을 이용하여 돌발성 난청 환자의 예후나,^{29,30)} 청신경초종 수술 환자에서 청력 보존이나 재발 여부,^{31,32)} 그리고 인공와우 수술 결과를 예측하기도 한다.³³⁾ 그리고 CNN 모델을 통해서 이내시경이나 안면 사진, CT/MRI 이미지를 분석하여 고막 천공이나 중이염, 외이기형, 신경초종 등을 자동 진단하는 연구들도 시행되었다.³⁴⁻³⁹⁾

비과 영역에서 머신러닝 연구

비과 영역에서도 머신러닝을 이용한 연구들이 활발히 진행 중이다. 초기에는 머신러닝 기법을 이용하여 만성 부비동염의 표현형(phenotype)을 분류하는 연구가 주된 흐름이었으나, 근래에는 CNN과 같은 딥러닝 기법을 이용한 연구들이 점차 늘고 있는 추세이다(Table 2).

만성부비동염 환자들에게 맞춤형 치료 전략을 구사하기 위해서는 환자들의 임상 소견에 따른 표현형뿐만 아니라, 분자 수준의 면역학적 특성을 반영한 내재형(endotype)까지 고려되어야 한다. 이를 위해서 다양한 바이오마커(biomarker)를 이용하여 만성부비동염을 분류하는 시도들이 있어 왔다. 이

Table 2. Summary of studies with machine learning application in rhinology

Author, year	Application	Machine learning algorithm	Outcome
Soler, et al., ⁴⁰⁾ 2016	CRS phenotyping using clinical findings	Unsupervised clustering analysis	5 Clusters with the most discriminating variables being age, productivity loss, and total SNOT-22
Lal, et al., ⁴¹⁾ 2018	CRS classification using SNOT-22 scores	Unsupervised clustering analysis	4 Clusters with severe, moderate with sinonasal symptoms, moderate without sinonasal symptoms, and mild
Tomassen, et al., ⁴²⁾ 2016	CRS endotyping using biomarkers (cytokines)	Unsupervised clustering analysis	10 Clusters, three were strongly associated with presence of polyposis and asthma, while another three were associated with presence of non-polyp disease and low incidence of asthma
Kim, et al., ⁴³⁾ 2019	CRSwNP classification using clinical findings	Two-step cluster analysis	6 Clusters, two were associated with asthma and four were not associated with asthma
Chowdhury, et al., ⁴⁵⁾ 2019	Classify osteomeatal complex occlusion on CT	Transfer learning with Google Inception-V3 CNN	Overall accuracy 85%, AUC 0.87
Huang, et al., ⁴⁶⁾ 2019	Differentiate the location of the anterior ethmoidal artery on CT	Transfer learning with Google Inception-V3 CNN	Overall accuracy 82.7%, AUC 0.86
Li, et al., ⁴⁸⁾ 2018	Detection and auto-segmentation of nasopharyngeal malignancies under endoscopic examination	CNN on the inception architecture	Overall accuracy 88.7%, AUC 0.94, DSC 0.78±0.24
Li, et al., ⁴⁹⁾ 2018	Automatic segmentation of MRI for nasopharyngeal carcinoma	CNN	DSC 0.89±0.05
Dimauro, et al., ⁵¹⁾ 2018	Identification and classification of nasal cytology	CNN	Sensitivity of cell identification 97%, accuracy of cell classification 99%

CRS: chronic rhinosinusitis, CRSwNP: CRS with nasal polyp, SNOT-22: Sino-nasal Outcome Test, CNN: convolutional neural network, AUC: area under the receiver operating characteristics, DSC: dice similarity coefficient

를 위해 머신러닝으로 빅데이터를 다루는 접근 방법을 통해 만성부비동염을 분류하고자 하는 시도들이 있어 왔는데, 대부분의 연구가 군집화 알고리즘(clustering algorithm)을 이용하였다. 군집화 알고리즘은 데이터를 특성의 유사함에 따라 일정 수의 군집으로 분류하는 알고리즘으로, 이때 입력된 데이터는 특성값만 존재하고 결과값이 존재하지 않으므로 머신러닝의 비지도 학습(unsupervised learning)에 해당한다. 머신러닝이 적용된 만성부비동염 분류 연구에서는 데이터의 특성에 따라 증상, 비용종, 흡연력, 천식 유무 등과 같은 임상 소견을 이용한 연구와 세포표지자(cell marker)나 사이토카인(cytokine)과 같은 바이오마커를 이용한 연구로 나눌 수 있다. 임상 소견을 데이터 특성으로 이용한 연구를 살펴보면, Soler 등⁴⁰⁾은 약물에 반응하지 않는 382명의 만성부비동염 환자에서 32개의 특성을 사용하여 군집화 분석을 시행했을 때, 대상 환자들이 5개의 군집으로 분류되었고, 나이, 생산성 저하, Sino-nasal Outcome Test(SNOT-22) 점수가 군집을 분류하는 데 가장 의미있는 특성이었다. 하지만 비용종 유무, 아토피, 천식, 아스피린 과민성 등과 같이 기존에 만성부비동염의 표현형을 구분하는 데 중요하게 여겨졌던 특성

들은 그 중요도가 낮은 것으로 보고하였다. Lal 등⁴¹⁾은 비용종을 동반하지 않은 146명의 환자에서 SNOT-22 점수에 따라 군집 분석을 시행하였는데, 4개의 군집(severe, moderate with sinonasal symptoms, moderate with psychological sleep symptoms, mild)으로 나눌 수 있으며, severe 군집의 경우에는, 비용종 환자와 마찬가지로 천식 및 조직내 호산구(tissue eosinophilia)와 강한 상관 관계를 보였다. 바이오마커들을 이용한 연구들을 보면, Tomassen 등⁴²⁾이 수술 후 조직에서 사이토카인을 이용한 군집 분석을 시행하였다. 대상 환자들은 10개의 군집으로 분류되었고, 이 중 3개의 군집이 비용종과 천식에 강한 상관관계를 보였지만 나머지 3개 군집에서는 반대의 결과를 보였다. Kim 등⁴³⁾은 비용종을 가지고 있는 375명의 환자에서 점막과 혈중의 호산구 수, 나이, CT 소견, 천식 유무 등 6개의 특성을 가지고 군집 분석을 시행하여, 6개의 군집으로 분류하였다. 이 중 2개의 군집에서는 천식과 관련이 되었지만, 나머지 4개의 군집에서는 상관성이 없었다. 이상의 군집 분석 연구들에서 임상 소견 특성을 이용한 연구와 바이오마커를 이용한 연구들의 결과를 비교해 보면, 바이오마커를 이용한 연구들에서는 비용종과 심한 증상 및

염증 사이에 강한 상관관계를 보인 반면, 임상 소견을 이용한 연구에서는 반대로 둘 간에 상관관계가 낮은 것으로 보여진다.

이러한 머신러닝 관련 연구들은 만성부비동염의 분류에서 기존 연구들과는 일부 상반되는 결과들을 보여 주고 있다. 하지만, 머신러닝 모델 자체의 특성상 몇 가지 제한점들이 있으므로 해석에 유의해야 할 필요가 있다. 우선, 연구 대상 환자들의 수가 일반적인 빅데이터에 비해 상대적으로 적다는 점이다. 일반적으로 이비인후과, 특히 비과 영역의 임상 연구에서 빅데이터로 부를 수 있을 만큼 충분한 데이터를 모으는 것은 현실적으로 어려운 경우가 많기 때문에, 이들 연구들의 결과를 해석함에 있어 이중 오류(type II error)의 가능성을 항상 염두에 두어야 한다. 또한, 연구자가 대상 환자와 데이터의 특성을 어떻게 선택하는가에 따라 모델의 성능 뿐 아니라 결과에도 영향을 끼칠 수 있음을 이해해야 한다.⁴⁴⁾

비과영역에서 머신러닝을 이용한 다른 분야는 여러 딥러닝 기법을 이용하여 영상 이미지를 자동 분류해 내는 연구들이다. 최근 Chowdhury 등⁴⁵⁾은 CNN 알고리즘을 이용하여 부비동 CT에서 osteomeatal complex의 폐쇄 여부를 자동으로 분류해 내는 딥러닝 모델을 보고하였다. 이 연구에서는 만성 부비동염 환자 239명의 부비동 CT에서 추출한 956개의 관상(coronal) 이미지에 osteomeatal complex의 폐쇄 여부를 labelling 한 다음, 기존에 이미 128만 장이 넘는 이미지들로 학습되어 있는 Google의 Inception-V3 CNN 모델을 통해 전이 학습을 시행하였다. 그 결과, osteomeatal complex의 폐쇄 여부 진단에서 85%[95% confidence interval(CI), 78~92%]의 정확도를 보였고 area under the receiver operating characteristics curve(AUC)값은 0.87이었다. Huang 등⁴⁶⁾은 675개의 부비동 CT 이미지를 Inception V3 모델로 전이 학습시킨 결과, CT상에서 전사골동맥(anterior ethmoidal artery)의 위치를 82.7%의 정확도와 AUC 0.86으로 분류해 낼 수 있었다. 이밖에 최근에는 690명의 환자에서 시행된 부비동 CT를 CNN 모델로 학습시켜서, CT상에서 부비동 부분만을 분할하여 부비동의 전체 부피와 혼탁 정도를 자동으로 계산하는 모델이 보고되기도 하였다.⁴⁷⁾

딥러닝 기법을 이용하여 비인두암을 진단하고자 하는 연구들도 시행되었다. Li 등⁴⁸⁾은 27536개의 비인두 내시경 이미지를 CNN 모델을 통해 학습시켰고, 비인두암 진단에서 88.0%(95% CI, 86.1~89.6%)의 정확도를 보여 전문의(80.5%)나 전공의(72.8%)보다 우수한 결과를 나타내었다. 또한 이 딥러닝 모델이 비인두암의 병변 부위를 분할하는 데에도 효과적이라고 보고하였다. Li 등⁴⁹⁾은 MRI 이미지를 CNN 모델에 학습시켜 비인두암의 병변을 자동 분할하는 모델을 만들었고, 이 모델의 dice similarity coefficient(DSC)는 0.89 ± 0.05

로 높은 정확성을 보고하였다. Liang 등⁵⁰⁾도 비인두암 환자에서 방사선 치료 범위를 결정하기 위해 두경부 CT에서 주요 장기를 자동 분할하는 CNN 모델을 보고하였는데, 뇌간, 안구, 수정체, 후두, 구강, 척수, 이하선, 하악, 유양돌기 등 대부분의 장기에서 DSC 0.85 이상의 정확성을 보였다.

최근에는 비과 영역의 병리 분야에서도 딥러닝 기법을 활용한 시도가 있었다. Dimauro 등⁵¹⁾은 디지털화된 nasal cytology 이미지에서 CNN 모델을 이용하여 세포를 자동 분리한 다음, 세포들의 종류까지 자동으로 구분하는 연구를 시행하였는데, 세포를 분리하는데 97% 민감도, 세포 종류 진단에는 99% 정도의 정확성을 보였다.

이 밖에 안면 성형 분야에서도 머신러닝 기법들이 적용되고 있다. Borsting 등⁵²⁾은 일반인에서 코 성형 시행 여부를 판단하기 위해 22686장의 코 성형 전후 안면 사진을 CNN 모델에서 학습시켰고, 안면 사진만으로 85%에서 코 성형 여부를 정확히 예측할 수 있었다. Dorfman 등⁵³⁾은 안면 사진을 통해 나이를 예측하는 CNN 알고리즘(Microsoft Azure Face API)에 100명의 코 성형 환자들의 수술 전후 사진을 전이 학습시킨 결과, 코 성형 수술 후 평균 약 3.1년 정도 실제 나이보다 젊게 보인다고 보고하기도 하였다.

결론

의료 분야에서 머신러닝이나 딥러닝과 같은 인공지능의 활용은 이제 큰 흐름이 되었다. 빅데이터와 정밀 의학 시대에 머신러닝의 활용은 점차 늘어날 것이고, 머신러닝 알고리즘의 개발에서 가장 중요한 부분인 양질의 데이터를 충분히 제공하기 위해서 임상 의사들의 역할도 점차 커질 것이다. 이비인후과 의사 역시 임상에서 머신러닝을 활용하기 위해서는 머신러닝 알고리즘의 기본 개념을 알고 데이터 과학자들과 소통할 수 있는 능력을 길러야 할 것이다.

Acknowledgments

None.

ORCID

Myeong Sang Yu <https://orcid.org/0000-0002-4529-0254>

REFERENCES

- 1) Bur AM, Shew M, New J. Artificial intelligence for the otolaryngologist: A state of the art review. *Otolaryngol Head Neck Surg* 2019;160(4):603-11.
- 2) Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods* 2018;15(4):233-4.
- 3) McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 1943;5(4):115-33.
- 4) Rosenblatt F. The perceptron: A probabilistic model for information

- storage and organization in the brain. *Psychol Rev* 1958;65(6):386-408.
- 5) Griewank A. Who invented the reverse mode of differentiation? *Doc Math Extra Vol. Optimization Stories* 2012:389-400.
- 6) Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. *ICML* 2010.
- 7) Dahl GE, Sainath TN, Hinton GE. Improving deep neural networks for LVCSR using rectified linear units and dropout. *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*; 2013 May 26-31, Vancouver, Canada.
- 8) Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017;318(22):2199-210.
- 9) Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;316(22):2402-10.
- 10) Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542(7639):115-8.
- 11) Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9(8):1735-80.
- 12) Lee HC, Ryu HG, Chung EJ, Jung CW. Prediction of bispectral index during target-controlled infusion of propofol and remifentanyl: A deep learning approach. *Anesthesiology* 2018;128(3):492-501.
- 13) Park SH. Artificial intelligence in medicine: Beginner's guide. *J Korean Soc Radiol* 2018;78(5):301-8.
- 14) van der Veen J, Willems S, Deschuymer S, Robben D, Crijns W, Maes F, et al. Benefits of deep learning for delineation of organs at risk in head and neck cancer. *Radiother Oncol* 2019;138:68-74.
- 15) Huang B, Chen Z, Wu PM, Ye Y, Feng ST, Wong CO, et al. Fully automated delineation of gross tumor volume for head and neck cancer on PET-CT using deep learning: A dual-center study. *Contrast Media Mol Imaging* 2018;2018:8923028.
- 16) Ren J, Jing X, Wang J, Ren X, Xu Y, Yang Q, et al. Automatic recognition of laryngoscopic images using a deep-learning technique. *Laryngoscope* 2020.
- 17) Mascharak S, Baird BJ, Holsinger FC. Detecting oropharyngeal carcinoma using multispectral, narrow-band imaging and machine learning. *Laryngoscope* 2018;128(11):2514-20.
- 18) Tamashiro A, Yoshio T, Ishiyama A, Tsuchida T, Hijikata K, Yoshimizu S, et al. Artificial intelligence-based detection of pharyngeal cancer using convolutional neural networks. *Dig Endosc* 2020.
- 19) Xiong H, Lin P, Yu JG, Ye J, Xiao L, Tao Y, et al. Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images. *EBioMedicine* 2019;48:92-9.
- 20) Halicek M, Little JV, Wang X, Chen AY, Fei B. Optical biopsy of head and neck cancer using hyperspectral imaging and convolutional neural networks. *J Biomed Opt* 2019;24(3):1-9.
- 21) Karadaghy OA, Shew M, New J, Bur AM. Development and assessment of a machine learning model to help predict survival among patients with oral squamous cell carcinoma. *JAMA Otolaryngol Head Neck Surg* 2019;145(12):1115-20.
- 22) Formeister EJ, Baum R, Knott PD, Seth R, Ha P, Ryan W, et al. Machine learning for predicting complications in head and neck microvascular free tissue transfer. *Laryngoscope* 2020.
- 23) Fujima N, Shimizu Y, Yoshida D, Kano S, Mizumachi T, Homma A, et al. Machine-learning-based prediction of treatment outcomes using MR imaging-derived quantitative tumor information in patients with sinonasal squamous cell carcinomas: A preliminary study. *Cancers (Basel)* 2019;11(6):800.
- 24) Mermod M, Jourdan EF, Gupta R, Bongiovanni M, Tolstonog G, Simon C, et al. Development and validation of a multivariable prediction model for the identification of occult lymph node metastasis in oral squamous cell carcinoma. *Head Neck* 2020;42(8):1811-20.
- 25) Alabi RO, Elmusrati M, Sawazaki-Calone I, Kowalski LP, Haglund C, Coletta RD, et al. Machine learning application for prediction of locoregional recurrences in early oral tongue cancer: A Web-based prognostic tool. *Virchows Arch* 2019;475(4):489-97.
- 26) Tighe D, Lewis-Morris T, Freitas A. Machine learning methods applied to audit of surgical outcomes after treatment for cancer of the head and neck. *Br J Oral Maxillofac Surg* 2019;57(8):771-7.
- 27) Shew M, New J, Bur AM. Machine learning to predict delays in adjuvant radiation following surgery for head and neck cancer. *Otolaryngol Head Neck Surg* 2019;160(6):1058-64.
- 28) Bur AM, Holcomb A, Goodwin S, Woodroof J, Karadaghy O, Shnyder Y, et al. Machine learning to predict occult nodal metastasis in early oral squamous cell carcinoma. *Oral Oncol* 2019;92:20-5.
- 29) Bing D, Ying J, Miao J, Lan L, Wang D, Zhao L, et al. Predicting the hearing outcome in sudden sensorineural hearing loss via machine learning models. *Clin Otolaryngol* 2018;43(3):868-74.
- 30) Park KV, Oh KH, Jeong YJ, Rhee J, Han MS, Han SW, et al. Machine learning models for predicting hearing prognosis in unilateral idiopathic sudden sensorineural hearing loss. *Clin Exp Otorhinolaryngol* 2020;13(2):148-56.
- 31) Cha D, Shin SH, Kim SH, Choi JY, Moon IS. Machine learning approach for prediction of hearing preservation in vestibular schwannoma surgery. *Sci Rep* 2020;10(1):7136.
- 32) Abouzari M, Goshtasbi K, Sarna B, Khosravi P, Reutershan T, Mostaghni N, et al. Prediction of vestibular schwannoma recurrence using artificial neural network. *Laryngoscope Investig Otolaryngol* 2020;5(2):278-85.
- 33) Kim H, Kang WS, Park HJ, Lee JY, Park JW, Kim Y, et al. Cochlear implantation in postlingually deaf adults is time-sensitive towards positive outcome: Prediction using advanced machine learning techniques. *Sci Rep* 2018;8(1):18004.
- 34) Livingstone D, Talai AS, Chau J, Forkert ND. Building an otoscopic screening prototype tool using deep learning. *J Otolaryngol Head Neck Surg* 2019;48(1):66.
- 35) Wang YM, Li Y, Cheng YS, He ZY, Yang JM, Xu JH, et al. Deep learning in automated region proposal and diagnosis of chronic otitis media based on computed tomography. *Ear Hear* 2020;41(3):669-77.
- 36) Hallac RR, Lee J, Pressler M, Seaward JR, Kane AA. Identifying ear abnormality from 2D photographs using convolutional neural networks. *Sci Rep* 2019;9(1):18198.
- 37) Livingstone D, Chau J. Oscopic diagnosis using computer vision: An automated machine learning approach. *Laryngoscope* 2020;130(6):1408-13.
- 38) Habib AR, Wong E, Sacks R, Singh N. Artificial intelligence to detect tympanic membrane perforations. *J Laryngol Otol* 2020;134(4):311-5.
- 39) George-Jones NA, Wang K, Wang J, Hunter JB. Automated detection of vestibular schwannoma growth using a two-dimensional U-Net convolutional neural network. *Laryngoscope* 2020.
- 40) Soler ZM, Hyer JM, Rudmik L, Ramakrishnan V, Smith TL, Schlosser RJ. Cluster analysis and prediction of treatment outcomes for chronic rhinosinusitis. *J Allergy Clin Immunol* 2016;137(4):1054-62.
- 41) Lal D, Hopkins C, Divekar RD. SNOT-22-based clusters in chronic rhinosinusitis without nasal polyposis exhibit distinct endotypic and prognostic differences. *Int Forum Allergy Rhinol* 2018;8(7):797-805.
- 42) Tomassen P, Vandeplas G, Van Zele T, Cardell LO, Arebro J, Olze H, et al. Inflammatory endotypes of chronic rhinosinusitis based on cluster analysis of biomarkers. *J Allergy Clin Immunol* 2016;

- 137(5):1449-56.e4.
- 43) Kim JW, Huh G, Rhee CS, Lee CH, Lee J, Chung JH, et al. Unsupervised cluster analysis of chronic rhinosinusitis with nasal polyp using routinely available clinical markers and its implication in treatment outcomes. *Int Forum Allergy Rhinol* 2019;9(1):79-86.
 - 44) Walker A, Surda P. Unsupervised learning techniques for the investigation of chronic rhinosinusitis. *Ann Otol Rhinol Laryngol* 2019;128(12):1170-6.
 - 45) Chowdhury NI, Smith TL, Chandra RK, Turner JH. Automated classification of osteomeatal complex inflammation on computed tomography using convolutional neural networks. *Int Forum Allergy Rhinol* 2019;9(1):46-52.
 - 46) Huang J, Habib AR, Mendis D, Chong J, Smith M, Duvnjak M, et al. An artificial intelligence algorithm that differentiates anterior ethmoidal artery location on sinus computed tomography scans. *J Laryngol Otol* 2020;134(1):52-5.
 - 47) Humphries SM, Centeno JP, Notary AM, Gerow J, Cicchetti G, Katial RK, et al. Volumetric assessment of paranasal sinus opacification on computed tomography can be automated using a convolutional neural network. *Int Forum Allergy Rhinol* 2020.
 - 48) Li C, Jing B, Ke L, Li B, Xia W, He C, et al. Development and validation of an endoscopic images-based deep learning model for detection with nasopharyngeal malignancies. *Cancer Commun (Lond)* 2018;38(1):59.
 - 49) Li Q, Xu Y, Chen Z, Liu D, Feng ST, Law M, et al. Tumor segmentation in contrast-enhanced magnetic resonance imaging for nasopharyngeal carcinoma: Deep learning with convolutional neural network. *Biomed Res Int* 2018;2018:9128527.
 - 50) Liang S, Tang F, Huang X, Yang K, Zhong T, Hu R, et al. Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning. *Eur Radiol* 2019;29(4):1961-7.
 - 51) Dimauro G, Ciprandi G, Deperte F, Girardi F, Ladisa E, Latrofa S, et al. Nasal cytology with deep learning techniques. *Int J Med Inform* 2019;122:13-9.
 - 52) Borsting E, DeSimone R, Ascha M, Ascha M. Applied deep learning in plastic surgery: Classifying rhinoplasty with a mobile app. *J Craniofac Surg* 2020;31(1):102-6.
 - 53) Dorfman R, Chang I, Saadat S, Roostaeian J. Making the subjective objective: Machine learning and rhinoplasty. *Aesthet Surg J* 2020;40(5):493-8.

정답 및 해설

답 ⑤

해설 위축성 비염(atrophic rhinitis)는 비강내 점막의 위축과 가피의 과도한 형성이 특징인 질환이며 악취를 동반하는 경우 취비증 등의 병명으로도 불린다. 주로 사회경제적 상태가 낮은 사람에게 잘 생기는 편이며, 남자에 비해 여자에게 생기는 빈도가 5.6배 정도 높으며, 비강에서 세균배양검사를 하면 *Klebsiella ozaenae*가 대부분 검출된다. 진단은 병력과 진찰만으로 가능하다.

전신증상이나 부비동염 등을 동반한 경우가 아니면 전신적 항생제 투여는 의미가 없으며, 비수술적 치료로는 선행요인의 해결, 가피제거, 비강세척을 통한 청결유지, 보습 등이 있다.

참고 문헌: 대한비과학회. 최신 임상비과학. 파주: 군자출판사;2017. p.238-41.

대한이비인후과학회. 이비인후과학. 서울: 일조각;2005. p.1113-4.