臨床耳鼻：第 28 卷　第 2 號　2017
J Clinical Otolaryngol 2017;28:320-323

왓슨 포 온콜리지란?

부산대학교 의학전문대학원 혈액종양내과학교실
김효정 · 최영진

What is Watson for Oncology?

Hyojeong Kim, MD and Young Jin Choi, MD
Division of Hemato-Oncology, Pusan National University College of Medicine, Busan, Korea

서　론

2016년 전국민의 관심을 끌었던 알파고와 이세돌의 바둑 대결을 지켜 보면서 의료계에서도 인공 지능(AI)에 대한 기대와 우려에 대한 다양한 시각이 나오고 있다. IBM의 인공지능(AI) 암 치료 솔루션 '왓슨'(Watson for Oncology)이 2016년 12월 가천대 길병원에 도입된 이후 부산대병원, 건양대병원, 동산병원, 대가대병원, 조선대병원, 전남대병원에서도 도입되었고, 실제 진료 환경에 접목되어 시행되고 있다.

왓슨은 의사의 의료적 의사결정 과정을 돕는 임상 의사결정 지원 시스템(clinical decision support system)으로 2012년 3월 미국 메모리얼 슬론 케터링 암센터(MSKCC)에서 폐암, 유방암, 전립선암 등 암 진료에 처음 도입됐다.

왓슨이 가진 핵심 기술은 자연어 처리(natural language processing)와 기계 학습(machine learning)으로 인간의 언어를 이해할 수 있어 방대한 양의 데이터를 스스로 분석하고 학습해 복잡한 문제에 대한 답을 빠른 시간 안에 내놓는다. 왓슨은 현재 암 환자의 치료법을 권고하고(Watson for Oncology), 유전정보를 분석하며(Watson for Genomics), 의료영상을 판독하고(Watson imaging clinical review), 임상시험을 도울(Watson for Clinical Trial Matching)수 있다고 한다.

본 종설에서는 왓슨, 특히 현재까지 우리나라의 7개 병원에서 도입한 왓슨 포 온콜로지란 무엇이며 어느 수준까지 개발된 단계인지와 앞으로의 전망은 어떠할 지에 대해 살펴 보기로 한다.

교신저자 : 최영진, 49241 부산광역시 서구 구덕로 179
부산대학교 의학전문대학원 혈액종양내과학교실
전화 : (051) 240-7852 · 전송 : (051) 256-8330
E-mail : chyj@pusan.ac.kr

본　론

왓슨 포 온콜로지는 무엇인가?

왓슨은 IBM이 만든 인공지능 또는 그 인공지능이 탑재된 슈퍼컴퓨터로서 초대 IBM 경영자의 이름에서 유래한 명칭이라고 한다. 2011년 미국의 유명 퀴즈쇼 제퍼디에 출연하여 우승하면서 유명세를 타기 시작했다. 왓슨 포 온콜로지는 암 진단에 특화된 IBM 왓슨이며 메모리얼 슬로언 케터링(MSK) 암센터에서 기계 학습을 하여 의사들이 근거에 입각한(evidence-based) 치료를 할 수 있도록 지원하는 목적으로 개발됐다.

IBM에 따르면(Pubmed) 2015년 한 해 동안 전세계적으로 약 4만 4,000건에 달하는 종양학 논문이 의료 학술지에 발표 되었고 이는 매일 122개의 새로운 논문이 발표된다는 의미다. 폭발적으로 증가하는 의료지식은 그 양과 속도에 있어 인간의 능력으로 따라갈 수 있는 한계를 넘어서고 있다. 왓슨 포 온콜로지는 300개 이상의 의학 학술지, 200개 이상의 의학 교과서를 포함해 거의 1,500만 페이지에 달하는 의료 정보를 이미 학습했다고

한다.

왓슨 포 온콜리지는 클라우드 기반의 플랫폼으로 방대한 분량의 정형(structured) 및 비정형(unstructured) 데이터를 분석해, 의사들이 암환자들에게 데이터에 근거한 개별화된 치료 옵션을 제공할 수 있도록 지원한다. 의사들이 자연어 처리가 가능한 '왓슨 포 온콜로지'에 특정 환자의 진료 기록과 의료 데이터를 입력하면 출판된 연구 결과와 임상 가이드라인 등을 확인하여 선택 가능한 치료법을 권고 받을 수 있다. 이러한 권고는 초록색, 주황색, 빨간색의 3단계로 제시되는데, 초록색은 추천하는(recommended) 치료법, 주황색은 고려해 볼 수 있는(for consideration) 치료법이며 빨간색은 권고하지 않는(not recommended) 치료법이다.

왓슨 포 온콜로지의 현실적 성적

왓슨을 학습시킨 MSK 암 센터에서 2013년 미국 임상종양학회에 발표한 초록은 525개의 실제 폐암 환자 증례와 420개의 가상 증례를 학습시킨 결과로서 왓슨의 자연어 처리 능력과 기계 학습 능력에 대해 평가하였고, 동일 증례 데이터를 대상으로 반복 학습 시에 이 2가지 능력이 향상되는 것을 보여주었다.[1] 적합한 치료 방침을 내놓는 능력이 MSK 의료진의 판단을 기준으로 해서 40%에서 77%까지 상승한다는 결과를 발표한 것이다.

2014년에는 이를 확장하여 대장암, 직장암, 방광암, 췌장암, 신장암, 난소암, 자궁경부암, 자궁내막암에 대해 학습시킨 결과를 같은 학회에서 발표하였다.[2] 역시 반복적인 학습을 통해 90~100%에 달하는 정확도를 보였다는 결과를 보여주었지만 이는 동일 증례를 여러 차례 반복하여 테스트한 방식이기 때문에 새로운 환자에 대한 의사 결정에 도움이 필요한 실제 진료 현장에서 얼마나 의미가 있을 지에 대해서는 의문이 있다고 할 수 있다. 2014년 미국 임상종양학회에서는 MSK 암센터 의료진 6명이 왓슨 시스템을 사용해 본 경험에 대한 피드백을 발표하기도 했는데, 결론적으로 왓슨이 도움은 되지만 불필요한 환자 데이터 입력에 시간이 많이 소요되고 권고하는 치료법의 우선순위가 어떻게 정해지는지에 대한 정보가 부족하다는 것을 문제점으로 꼽았다.[3]

이는 필자가 직접 사용해 본 경험과 동일한 것으로 데이터 입력 문제의 경우 자연어 처리와 의료 기록 연동 능력을 향상시킬 수 있다면 어느 정도 해결할 수 있겠지만 후자의 경우가 문제이다. 기계 학습의 기본적인 원칙이 "garbage-in, garbage-out"이라고 하여 질이 나쁜 데이터로 학습을 시키면 나쁜 결과가 나온다는 것이다. 근거중심의학에서 우리는 증례 보고와 무작위 대조군 연구, 체계적 문헌 고찰의 결과를 동일한 가치로 취급하지 않는다. 우리가 증거의 수준이라고 부르는 이러한 요소 외에도 데이터의 질에 영향을 미치는 요소는 여러 가지가 있을 것이다. 훈련시킬 때 투입된 데이터의 질에 대한 평가를 어떻게 했는지가 결국 추천의 우선순위를 결정하는 기준이 되었을 것인데 이에 대한 정보 제시가 부족하다. 다시 말해 왓슨이 의사 결정을 수행한 결과만이 아니라 그 과정도 보여주기를 의사들이 원한다는 이야기이다.

2015년 미국 임상종양학회에서 MSK 암 센터에서는 왓슨과 관련하여 3편의 초록을 발표하였다. 발표된 MSK 암센터 연구 결과에 따르면 1기에서 3기 유방암의 경우 90% 이상, 폐암의 경우 50% 가량 MSK 치료 권장사항과 일치했다.[4,5] 20 증례의 폐암 환자를 대상으로 한 연구에서 MSK 치료 선택이 왓슨과 일치하지 않는 50% 중 각 25%가 왓슨의 "for consideration"과 "not recommended" 옵션에 해당되었다. 왓슨 비추천에 해당된 경우는 동반 질환이 있는 고령 환자들의 경우로서 아직 왓슨의 기계 학습이 이루어지지 않은 탓으로 생각되었다. 전이성 폐암 16 증례에서 이미 투여된 항암 치료법의 88%가 "recommended" 또는 "for consideration"에 포함되어 있었다.

또한 101개의 훈련용 전이성 유방암 증례를 대상으로 하여 MSK 암 센터 의료진이 다양한 환자 요소에 대해 어떻게 가중치를 부여하여 의사 결정을 하는지 기계학습을 시킨 내용도 발표되었는데, 학습 전 정확도 73.6%에서 학습 후 정확도 82.1%의 향상을 보고하였다.[6] 이 연구 결과는 기존의 방대한 양의 데이터를 분석할 때 데이터의 질 인식에 대한 의구심 문제를 우회할 수 있는 가능성을 보여주었다는 점에서 주목할 만하다. 훈련시킨 사람의 의사결정 모델을 그대로 모방할 수 있는 가능성을 보여준 것이다.

2016년 12월 인도의 Manipal병원이 유럽 종양학회 아시아 회의에서 초록으로 발표한 후향적 연구 결과에서는

유방암(638), 대장암(126), 직장암(124), 폐암(112) 환자 총 1,000명의 증례를 대상으로 하여 의료진과 왓슨의 치료 방침을 비교하였다.[7] 의료진의 치료 방침이 왓슨의 "recommended"에 해당하는 환자의 비율은 직장암에서 85%로 가장 높았고 폐암에서 17.8%로 가장 낮았다. 전체 환자를 대상으로 보았을 때"recommended"와 "for consideration"을 합한 비율의 일치도는 80%로 확인되었다.

올해인 2017년 미국 임상종양학회에서는 태국의 범룽랏 병원에서 폐암, 유방암, 위암 환자 92명의 후향적 분석과 119명의 전향적 분석이 포함된 연구 결과를 보고하였는데, 각 암종별로 91%, 76%, 78%의 일치도를 보였다.[8] 또한 우리나라 길병원에서 대장암 환자 340명과 위암 환자 185명을 대상으로 한 후향적 연구 결과에서 73%와 49%의 일치도를 보고하였다.[9] 위암 환자에서 일치율이 낮은 이유로는 특정 항암 치료가 치료 당시 보험 급여를 받지 못하는 상황이었던 점과 위암이 많이 발생하는 일본과 국내에서 활발하게 사용되는 항암제가 미국에서는 흔하게 사용되는 약제가 아닌 점을 제시하였다.

앞으로의 전망

앞에서 살펴 본 바와 같이 왓슨 포 온콜리지의 성적은 기본적으로 다학제적 협진을 포함하여 결정된 의료진의 치료 방침과의 일치도를 확인하는 것을 기본으로 확인되고 있다. 또한 2017년 현재까지 발표된 내용은 대부분이 후향적 연구로 초록으로만 보고되어 학술 논문으로 출판된 내용은 없는 실정이다. 전향적 연구의 확대와 논문 출판이 이루어짐으로써 그 유용성에 대한 근거가 더 확보되어야 할 것이다.

일단 의료진의 치료 방침과의 일치도가 높은 것이 성적의 기준이 된다면 의사가 왓슨을 사용하여 얻을 수 있는 추가적인 이득이 무엇인가에 대해 생각해 볼 필요가 있다. IBM에서는 암 전문 의료인력이 부족한 지역에서 의사들이 왓슨에 의지하여 진료를 볼 수 있을 것이라는 취지를 처음부터 이야기해 왔다. 그러나 우리나라와 같이 의료 접근성이 비교적 좋은 환경에서는 실효성이 낮은 이야기가 될 수 있을 것이다.

2016년 샌안토니오 유방암 심포지엄에서 인도의 마니팔 병원이 발표한 유방암 환자 638명을 대상으로 한 연구는 시간 단축에 관한 언급을 한다.[10,11] 수작업으로 진행할 때 의료진들이 치료 방침 결정에 걸리는 시간은 20분 가량으로 비슷한 증례에 익숙해지면서 12분 가량으로 단축되나 왓슨의 경우 데이터를 분석하고 치료 방침을 권고하는 데에 40초 밖에 걸리지 않았다는 것이다. 대부분의 종양내과 의사들이 각 환자의 치료 방침을 결정하는 데에 진료실 밖에서도 많은 시간을 투자하는 것이 사실이다. 짧은 진료 시간 안에 복잡한 의사 결정을 빠르게 내리기 어렵기 때문이다. 현재 우리나라의 진료 환경을 고려하였을 때 데이터 입력에 대한 부분이 개선된다면 왓슨 포 온콜리지의 실질적인 유용성은 이러한 부분에서 찾을 수도 있을 것이다.

결 론

1980년대 말부터 등장한 근거중심의학은 2000년대 초반부터 우리나라 의료계에서도 정착되기 시작하였다. 비판적인 시각도 있지만 그 한계를 감안하여 현재 우리는 의료 현장에서 근거중심의학을 바탕으로 진료를 수행하고 있다. 그러나 쏟아져 나오는 연구 결과들의 양과 속도는 인간의 한계를 시험하는 수준에 다다르고 있다. 왓슨 포 온콜리지는 몇 가지 개선점을 보완한다면 우리가 근거중심의학을 충실히 수행하는 데에 도움을 줄 수 있을 것이다.

중심 단어 : 인공 지능 · 왓슨 포 온콜로지.

REFERENCES

1) Bach P, Zauderer MG, Gucalp A, Epstein AS, Norton L, Seidman AD, *et al. Beyond Jeopardy!: Harnessing IBM's Watson to improve oncology decision making. Journal of Clinical Oncology 2013;31.*
2) Epstein AS, Zauderer MG, Gucalp A, Seidman AD, Caroline A, Fu J, *et al. Next steps for IBM Watson Oncology: Scalability to additional malignancies. Journal of Clinical Oncology 2014;32.*
3) Zauderer MG, Gucalp A, Epstein AS, Seidman AD, Caroline A, Granovsky S, *et al. Piloting IBM Watson Oncology within Memorial Sloan Kettering's regional network. Journal of Clinical Oncology 2014;32.*
4) Kris MG, Gucalp A, Epstein AS, Seidman AD, Fu J, Keesing J, *et al. Assessing the performance of Watson for oncology, a decision support system, using actual contemporary clinical cases. Journal of Clinical Oncology 2015;33.*

5) Seidman AD, Pilewskie ML, Robson ME, Kelvin JF, Zauderer MG, Epstein AS, *et al. Integration of multi-modality treatment planning for early stage breast cancer (BC) into Watson for Oncology, a Decision Support System: Seeing the forest and the trees. Journal of Clinical Oncology 2015;33.*
6) Fu J, Gucalp A, Zauderer MG, Epstein AS, Kris MG, Keesing J, *et al. Steps in developing Watson for Oncology, a decision support system to assist physicians choosing first-line metastatic breast cancer (MBC) therapies: Improved performance with machine learning. Journal of Clinical Oncology 2015;33.*
7) Somashekhar S, Kumar R, Kumar A, Patil P, Rauthan A. *Validation study to assess performance of IBM cognitive computing system Watson for oncology with Manipal multidisciplinary tumour board for 1000 consecutive cases: An Indian experience. Annals of Oncology 2016;27.*
8) Suwanvecho S, Suwanrusme H, Sangtian M, Norden AD, Urman A, Hicks A, *et al. Concordance assessment of a cognitive computing system in Thailand. Journal of Clinical Oncology 2017;35.*
9) Baek JH, Ahn SM, Urman A, Kim YS, Ahn HK, Won PS, *et al. Use of a cognitive computing system for treatment of colon and gastric cancer in South Korea. Journal of Clinical Oncology 2017;35.*
10) Somashekhar SP, Kumarc R, Rauthan A, Arun KR, Patil P, Ramya YE. *Double blinded validation study to assess performance of IBM artificial intelligence platform, Watson for oncology in comparison with Manipal multidisciplinary tumour board-First study of 638 breast cancer cases. Cancer Research 2017;77.*
11) Joo YH. *Application of big data for otolaryngologic diseases. J Clinical Otolaryngol 2016;27:405-7.*