

우수성과	바이오 마커를 활용한 암 유형 예측 및 주요 마커 선정을 위한 그래프 기반 딥러닝 모델 개발
연구자	김지희 (동국대학교 컴퓨터·과학 교수)

인공지능으로 췌장암 정밀진단 앞당긴다

“*Technology is nothing. What’s important is that you have a faith in people.*”
(기술은 아무것도 아니에요. 중요한 건 사람에 대한 믿음이지요.)

- 1994년 스티브 잡스 인터뷰 중

▲ 2010년 아이폰4를 발표하는 스티브 잡스 위키미디어, Matthew Yohe

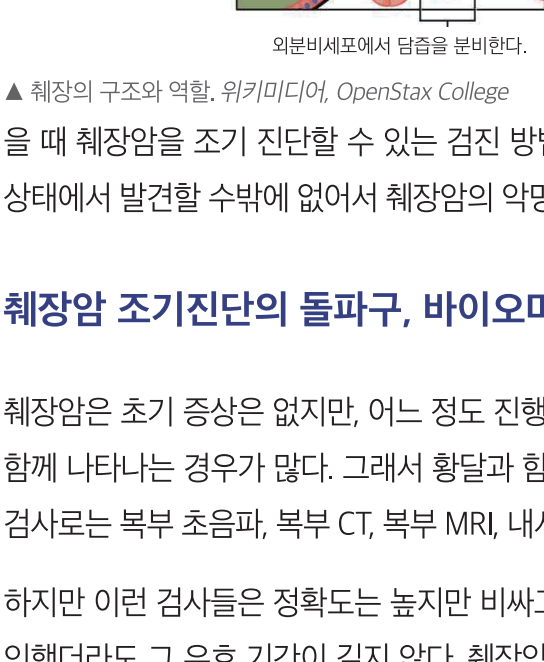
스티브 잡스, 췌장암으로 잠들다

2011년 10월 5일, 세계적인 기업가이자 혁신의 아이콘으로 불리던 스티브 잡스가 영면에 들었다. 당시 그의 나이 56세였다. 아이폰의 성공으로 세계적인 기업가이자 부호에 등극한 그를 죽음으로 이끈 건 8년 전 발병한 췌장암이었다. 2003년 10월 췌장암을 발견하고 종양 제거 수술과 치료, 전이로 인한 이식 수술 등을 받았지만 결국 그는 합병증으로 세상을 떠났다. 스티브 잡스의 죽음은 생존율을 낮기로 유명한 ‘악명 높은 암’으로 불리는 췌장암에 대해 다시 한번 관심을 불러일으키는 계기가 되었다.

췌장은 우리 몸에서 복부 안쪽, 위장 뒤쪽에 깊숙이 자리하고 있는 장기다. 15cm 정도의 길고 가는다란 모양을 하고 있으며, 우리 몸에서 소화를 돕는 외분비 기능과 혈당을 조절하는 내분비 기능을 맡고 있다. 이때 혈당 조절은 췌장의 랑게르한스섬에서 이뤄지는데, 여기서 나오는 호르몬이 인슐린과 글루카곤이다.

췌장암은 크게 외분비암과 내분비암으로 나누는데, 전체 췌장암의 90% 이상은 외분비샘에서 발생하는 선암으로 알려져 있다. 반면 내분비암은 5% 미만으로 드물게 발병하는데, 췌장의 랑게르한스섬 세포에서 발생하는 암은 ‘신경내분비종양(NET)’라고도 한다. 스티브 잡스가 앓았던 암이 바로 췌장암 중에서도 신경내분비종양이다.

췌장암의 악명이 높은 건 조기 발견이 어렵고, 그만큼 생존율이 상대적으로 낮기 때문이다. 초기에는 별다른 증상이 나타나지 않는다. 췌장 주변에 혈관이 많아 암을 발견했을 땐 다른 장기로 전이가 됐을 가능성이 높다. 우리나라 국가암정보센터 통계에 따르면 2018년에서 2022년까지 최근 5년간 암 종별 5년 상대생존율에서 췌장암이 현저하게 낮은 수치를 보였다. 상대생존율은 암 발생자와 동일한 연도, 성별, 연령의 일반인의 5년 기대생존율과 비교해 해당 암환자가 생존할 확률이다. 즉, 상대생존율이 100%라면 일반인의 생존율과 동일하다는 것을 뜻한다. 분석 결과 갑상선암이 101.1%로 가장 높았고, 전립선암 96.4%, 유방암 94.3%로 순으로 높게 나타났다. 이에 비해 폐암은 40.6%, 간암은 39.4%, 담낭 및 기타 담도암은 29.4%로 상대적으로 낮았다. 그리고 췌장암은 16.5%로 가장 낮은 수치를 보였다.



▲ 췌장의 구조와 역할. 위키미디어, OpenStax College

췌장에서 발견할 수밖에 없어서 췌장암의 악명은 계속되고 있다.

암은 진행 정도에 따라 ‘국한(Localized)’, ‘국소진행(Regional)’, ‘원격전이(Distant)’등으로 단계를 나눌 수 있다. 이중 암이 발생한 장기를 벗어나지 않는 국한 병기, 즉 초기에 발견하는 것이 생존에 유리하다. 실제로 통계에서도 국한 병기에선 진단했을 경우 유 갑상선암은 일반인과 동일한 생존율을 보였다. 이어 유방암 99.1%, 신장암 98.1%, 위암 97.4%, 폐암 79.8%, 간암 62.3%, 췌장암 46.6% 순이었다. 이 수치로만 보면 췌장암은 초기에 진단을 해도 생존율이 상대적으로 낮다. 하지만 암이 주위 장기나 림프절을 침범하는 국소 진행시에는 22.9%, 멀리 떨어진 부위로 퍼지는 원격 전이시에는 2.6%까지 생존율이 더욱 급격히 낮아진다.

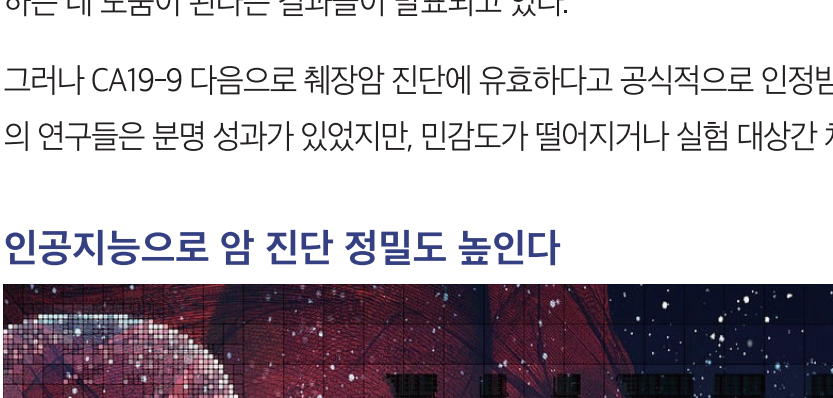
다행히 췌장암은 발병률이 매우 낮은 암이다. 우리나라 기준으로 췌장암의 발병률은 인구 1만명 당 1.5명이고, 평생 췌장암에 걸릴 확률도 1.6% 정도로 낮다. 하지만 아직은 별다른 증상이 없을 때 췌장암을 조기 진단할 수 있는 검진 방법이 딱히 없다. 결국 우연히 발견되는 게 아니라면 어느 정도 진행된 상태에서 발견할 수밖에 없어서 췌장암의 악명은 계속되고 있다.

췌장암 조기진단의 돌파구, 바이오마커

췌장암은 초기 증상은 없지만, 어느 정도 진행이 되면 복부 및 등 통증, 소화불량과 체중 감소, 황달 등 여러 증상이 함께 나타나는 경우가 많다. 그래서 황달과 함께 이런 증상들이 나타나면 병원에서 진단 검사를 받아야 한다. 진단 검사로는 복부 초음파, 복부 CT, 복부 MRI, 내시경 초음파 등이 대표적이다.

하지만 이런 검사들은 정확도는 높지만 비싸고 방사선 등의 위험도도 있다. 게다가 검사를 통해 암이 아닌 것을 확인했다더라도 그 유효 기간이 길지 않다. 췌장암은 검진 후 6개월 이내에도 빠르게 자라 퍼질 수 있는 공격적인 암이기 때문이다. 이러한 점 때문에 미국 보건복지부 산하 미국질병예방특별위원회(USPSTF)에서는 증상이 없는 사람은 이러한 검진을 받지 말도록 권고하고 있다.

그렇다면 안전하면서도 정확도 높은 조기진단 방법은 전혀 없는 것일까. 과학자들은 이에 대한 실마리를 혈액에서 찾고 있다. 바로 ‘바이오마커(biomarker)’다. 바이오마커는 인체 또는 그 생물체에서 측정될 수 있고, 결과 또는 질병의 발생에 영향을 미치거나 예측할 수 있는 모든 물질, 구조 또는 과정으로 정의된다. 암은 여러 가지 이유로 유전자에 변이가 생겼을 때, 세포가 비정상적으로 변하며 과다 증식해 일어나는 질환이다. 그래서 암세포가 자라기까지 혈액, 침, 소변 등에 관련한 대사 물질들이 흔적처럼 남게 된다. 이처럼 암 세포에서 분비되거나, 또는 암세포에 대한 반응으로 정상 세포에서 분비하는 특이적 물질을 ‘종양표지자’라고 한다. 그리고 췌장암에서 가장 널리 쓰이고 있는 바이오마커가 종양표지자 ‘CA19-9’다.



▲ AI 생성 이미지

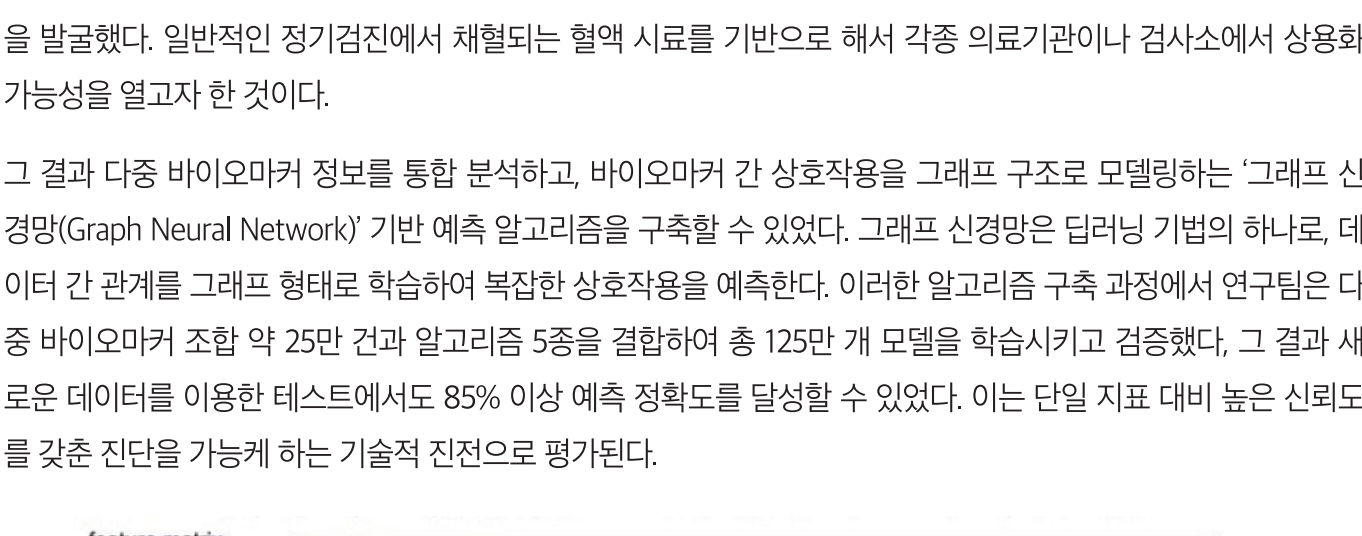
운 질환에서도 수치가 올라가기도 한다. 그래서 현재 CA19-9는 췌장암에 대한 진단보다는 치료의 경과나 예후 관찰, 재발 판정 등에 주로 활용된다.

CA19-9는 당지질 분자로 이뤄진 일종의 항원으로, 1979년 대장암 세포주에서 처음 발견되었다. 이후 대장암과 췌장암 환자의 혈청에서도 발견되며 진단검사로 개발되었는데, 현재는 주로 췌장암 환자들에게 적용되고 있다. 췌장염이나 췌장 낭종, 양성 담낭성 질환 등 여러 양성 위장 질환에서 과발현되는 특징이 있으며, 특히 종양 질환에서 혈장 수치가 급격히 증가한다. 하지만 CA19-9는 갑상선 질환이나 류마티스, 염증성 장질환 등 췌장암이 아닌 다른 가벼운 질환에서도 수치가 올라가기도 한다. 그래서 현재 CA19-9는 췌장암에 대한 진단보다는 치료의 경과나 예후 관찰, 재발 판정 등에 주로 활용된다.

이러한 한계를 극복하기 위해 ‘다중(multi) 바이오마커’에 대한 연구가 꾸준히 이뤄지고 있다. 하나가 아닌 여러 개의 바이오마커를 함께 사용해 정확도를 높이는 것이다. 또한 당류 외에 DNA, RNA와 같은 유전물질을 바이오마커로 활용하는 연구도 활발하다. 세포는 기능을 발현하고 신호를 전달하기 위해 작은 소포체나 입자를 분비하는데, 이 입자 속에는 DNA, mRNA, miRNA와 다양한 단백질들이 들어 있다. RNA는 DNA로부터 유전정보에 따라 단백질이 합성되는 과정에서 유전자 발현을 조절하는 역할을 하는 물질이다. 이 과정에서 mRNA와 miRNA가 중요한 역할을 한다는 사실이 밝혀졌고, 이 때문에 암과 같은 질환의 진단뿐만 아니라 치료에 이르기까지 중요한 생물학적 지표로 연구되고 있다. 실제로 여러 연구에서 mRNA·miRNA 바이오마커가 췌장암 환자의 건강한 대조군을 구별하는 데 도움이 된다는 결과들이 발표되고 있다.

그러나 CA19-9 다음으로 췌장암 진단에 유효하다고 공식적으로 인정받는 바이오마커는 아직 나오지 않았다. 그동안의 연구들은 분명 성과가 있었지만, 민감도가 떨어지거나 실험 대상간 차이가 나는 등의 한계를 보였기 때문이다.

인공지능으로 암 진단 정밀도 높인다



▲ Pixabay

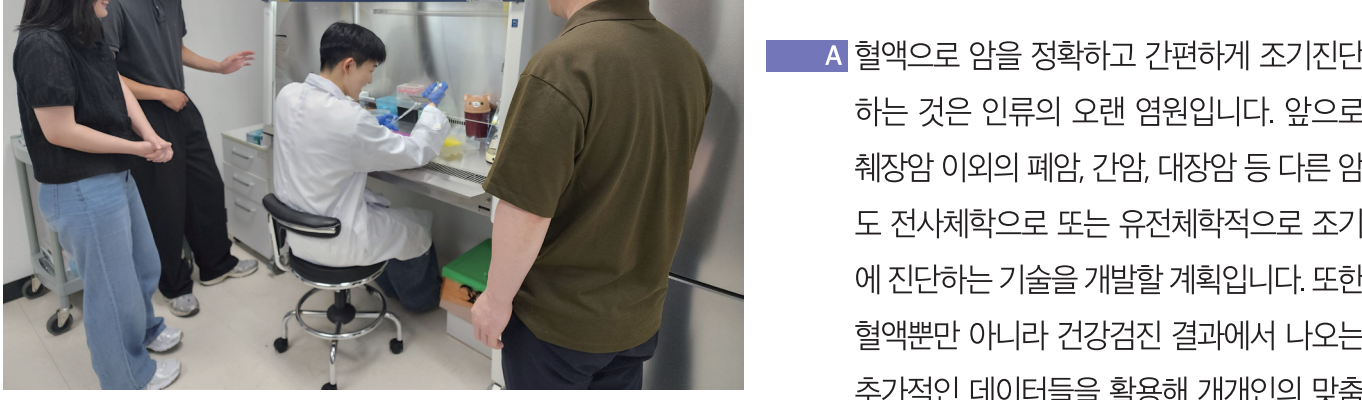
여러 물질의 조합이 가능한 다중 바이오마커와 유전체와 전사체, 단백질 등을 다루는 ‘오믹스(Omics)’ 기술의 등장으로 바이오마커 연구에서 데이터의 중요성은 갈수록 커지고 있다. 그런데 이렇게 방대한 데이터를 기존 방법으로 분석하는 건 불가능하다. 이때 새로운 돌파구를 마련해 준 것이 바로 인공지능(AI)이다. 딥러닝과 머신러닝 등 인공지능을 이용하면 대규모 데이터 세트에서 패턴을 찾아 분석할 수 있다. 특히 유전자 요인뿐만 아니라 생활방식, 환경 등 다양한 요소를 반영할 수 있어 개인별 맞춤 의료에 필요한 아이디어와 정보도 얻을 수 있다.

국내에서도 인공지능을 활용해 췌장암 고위험군을 정밀하게 예측할 수 있는 연구가 주목을 받고 있다. 바이오마커를 활용해 암 유형을 예측하고 주요 마커를 선정할 수 있는 그래프 기반 딥러닝 모델을 개발한 동국대학교 김지희 교수팀이 주인공이다.

연구팀은 췌장암의 표준화된 바이오마커인 CA19-9의 한계를 극복하기 위해서는 환자의 혈액을 분석해 병의 유형과 위험도를 정밀하게 예측할 수 있는 알고리즘 개발이 필요하다고 보았다. 즉, 바이오마커를 활용해 암 유형을 예측할 수 있는 그래프 기반 모델을 개발하고, 나아가 주요 바이오마커 선정을 통해 모델을 검증하고자 한 것이다.

이를 위해 연구팀은 우선 문헌 검토를 통해 55개의 후보 mRNA 바이오마커를 선정했다. 이후 선별과정을 거쳐 최종적으로 19개 바이오마커에 대해 췌장암을 특징하는 최적의 마커 패널과 알고리즘을 탐색했다. 19개 마커로 만들 수 있는 모든 조합의 경우의 수를 시도하고, 알고리즘별로 일일이 최적의 패널 조합을 찾은 것이다. 이때 췌장암 양성 판정을 받은 환자의 혈액 시료를 사용하여 다중 바이오마커의 발현량을 비교해 높은 예측도를 보이는 알고리즘을 발굴했다. 일반적인 정기검진에서 채혈되는 혈액 시료를 기반으로 해서 각종 의료기관이나 검사소에서 상용화 가능성을 열거한 것이다.

그 결과 다중 바이오마커 정보를 통합 분석하고, 바이오마커 간 상호작용을 그래프 구조로 모델링하는 ‘그래프 신경망(Graph Neural Network)’ 기반 예측 알고리즘을 구축할 수 있었다. 그래프 신경망은 딥러닝 기법의 하나로, 데이터 간 관계를 그래프 형태로 학습하여 복잡한 상호작용을 예측한다. 이러한 알고리즘 구축 과정에서 연구팀은 다중 바이오마커 조합 약 25만 건과 알고리즘 5종을 결합하여 총 125만 개 모델을 학습시키고 검증했다. 그 결과 새로운 데이터를 이용한 테스트에서도 85% 이상 예측 정확도를 달성할 수 있었다. 이는 단일 지표 대비 높은 신뢰도를 갖춘 진단을 가능케 하는 기술적 진전으로 평가된다.



▲ 그래프 신경망 기반 딥러닝 및 데이터 구성에 대한 흐름도. 동국대학교

이번 연구는 딥러닝 기반 알고리즘이 바이오마커 사이의 관계를 구조적으로 학습하도록 해서 예측 정확도와 해석 가능성을 동시에 높였다는 점이 핵심 성과라고 할 수 있다. 기존의 단일 바이오마커 기반이나 선형 분석 중심 진단 연구가 생체 신호의 상호작용을 충분히 설명하지 못했던 한계를 극복하는 계기를 마련한 것이다. 또한 다양한 마커 조합들을 비교해 성능을 정량적으로 검증함으로써 AI 기반 정밀진단 기술의 표준화와 재현성 확보에 기여한 점도 빼놓을 수 없다.

2025년 12월 국제학술지 ‘네이처’에서는 2026년 주목할 만한 과학 사건을 발표했다. 그중에는 복잡한 과학 연구 과정을 수행하는 AI 에이전트의 확장도, 50가지 유형의 암을 미리 감지할 수 있는 단일 혈액 검사에 대한 임상 시험 결과 발표가 들어 있었다. 과학계에서 인공지능과 진단 의학 분야가 그만큼 주목받고 있다는 걸 알 수 있는 대목이다. 주목받는 두 분야가 결합된 김지희 교수팀의 연구가 앞으로 어떤 성과로 이어질지 더욱 기대된다.

《#》미니 인터뷰

Q 어떤 계기로 이 주제를 연구하기 시작하셨는지 궁금합니다.

A 제가 운영하는 지식과학연구센터의 테마는 ‘Care(케어)’로, 사람들의 생활케어와 건강케어에 도움이 되는 인공지능 기술들을 개발하고 있습니다. 마침 췌장암 조기진단 시스템에 대하여 고민하고 있던 (췌췌벡)바이오와 공동연구를 시작하게 되었고, 자연스럽게 산학협력 모델로 이어지게 되면서 국가과제 공동수행, 공동 특허 출원, 기술이전 등 성과를 낼 수 있었습니다.



▲ 동국대학교 컴퓨터·과학과 김지희 교수. 동국대학교

Q 현재 이번 성과와 관련해 후속 연구도 진행하고 계신가요?

A 저희가 개발한 췌장암 조기진단시스템은 혈액을 이용한 비침습적인 검사 체계입니다. 혈액 시료를 전사체학적으로 취급하는 방법에도 여러 가지가 있을 수 있고, 병원 등 현장에서 더욱 간편하게 검사 가능한 방법도 개발할 수 있을 것 같아 관련 연구를 계속하고 있습니다. 또한 이번 연구에서 확립한 방법론을 췌장암뿐 아니라 다른 암이나 질환에도 적용 가능한 범용 진단 프레임워크로 확장하는 연구도 하고 있습니다.



▲ 연구현장의 모습. 동국대학교

Q 진단 의학은 영역이 앞으로 크게 확장될 것으로 보입니다. 앞으로 도전에 보고 싶은 연구 주제가 궁금합니다.

A 혈액으로 암을 정확하고 간편하게 조기진단하는 것은 인류의 오랜 염원입니다. 앞으로 췌장암 이외의 폐암, 간암, 대장암 등 다른 암도 전사체학으로 또는 유전체학적으로 초기에 진단하는 기술을 개발할 계획입니다. 또한 혈액뿐만 아니라 건강검진 결과에서 나오는 추가적인 데이터들을 활용해 개개인의 맞춤형 건강케어 시스템을 개발하고자 합니다.

Q 인공지능과 바이오가 융합된 연구를 하고 계신데요, 이 분야에 거는 대한 기대와 함께 고민도 있으신지 궁금합니다.

A 생성형 AI 등 인공지능 기술이 발전하면서 헬스케어 분야에도 많은 변화를 일으키고 있습니다. 바이오 및 건강 분야에 특화된 생성형 모델을 활용하여 데이터를 효과적으로 증강하는 등 다양한 시도가 진행되고 있는데, 앞으로 더 많은 기술적 발전이 있을 거라고 생각합니다.

특히 암 진단 분야는 인체 유래물을 연구 대상으로 한다는 점에서 막대한 비용과 시간이 소요됩니다. 하지만 아직 우리나라는 미국, 중국, 유럽 등에 비해 자원 투입이 부족한 실정입니다. 또 의료나 임상시험 관련 규제도 더 많은 편입니다. 연구를 위해 필요한 의료 데이터의 접근성에 대한 심층적인 논의가 필요하다고 생각합니다.

Q 이번 연구를 진행하는 과정에서 가장 결정적인 순간을 꼽는다면?

A 공동연구자 보니 다른 분야에 대한 이해도를 높이고 커뮤니케이션 방식을 확립하는 데에 처음에는 시간이 오래 걸렸습니다. 또한 양질의 데이터가 효과적인 인공지능 모델 개발의 가장 중요한 요소인데, 충분한 췌장암 샘플 데이터를 수집하는 데에 어려움이 있었습니다.

우선 여러 차례 세미나를 통해 소통의 장을 마련하려고 노력하였습니다. 동시에 실험에 필요한 데이터를 확보하기 위해 분당 서울대병원과 협력을 진행했습니다. 결국 양질의 데이터를 확보할 수 있었고, 이를 통하여 효과적인 조기진단 방법론을 구체화할 수 있었습니다.

이 모든 노력이 모여져 췌장암 조기진단 시스템 분자진단 데이터를 분석하는 머신러닝 모델을 만드는 과정과 관련한 표준 절차를 확립할 수 있었고, 특허 출원 및 등록까지 달성하였습니다. 이때가 가장 기억이 남습니다.

▲ 이번 연구에 함께한 연구팀의 모습. 동국대학교

글 고선아(길빛빛콘텐츠연구소 대표) 외
기획 한국과학기술기획평가원(김주희 외)