**Explainability in deep learning**

**1).전문가 지식과 딥러닝 모델의 차이 있다는 연구**

Most of the cases in which only the AI system identified cancer were invasive (Extended Data Table 5). On the other hand, cases in which only the reader identified cancer were split more evenly between in situ and invasive.

Notably, the additional cancers identified by the AI system tended to be invasive rather than in situ disease.

* McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., ... & Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*(7788), 89-94.
* 임상의와 모델이 포착하는 에러의 경우가 달랐다.
  + 서로 다른 것을 포착한다.

Given its recent groundbreaking success, AI holds the promise to significantly later the way we diagnose and stratify cancer in pathology

* Pati, P., Jaume, G., Foncubierta-Rodriguez, A., Feroce, F., Scognamiglio, G., Anniciello, A. M., ... & Gabrani, M. (2022). Graph Representation Learning and Explainability in Breast Cancer Pathology: Bridging the Gap between AI and Pathology Practice. *Artificial Intelligence Applications In Human Pathology*, 243.

+ Adversarial attack 연구들

* 사람이 보는 것과, 모델이 판단하는 것 사이에는 차이가 있음을 증명?

**3) 업무 만족도 높으면 워크 로드 감소에 도움이 되나?**

* 업무 만족도는 직원의 업무 성과와 긍정적이고 유의한 관련이 있다.
  + Siengthai, S., & Pila-Ngarm, P. (2016, August). The interaction effect of job redesign and job satisfaction on employee performance. In *Evidence-based HRM: a Global Forum for Empirical Scholarship*. Emerald Group Publishing Limited.
* 직무 만족도는 업무 과중과 업무 성과와의 관계를 완화해주지 못한다.
  + Jalal, R. N. U. D., & Zaheer, M. A. (2017). Does job satisfaction mediate the relationship of workload, remuneration and psychological reward with job performance. *International Journal of Academic Research in Business and Social Sciences*, *7*(9), 64-79.
* 업무 만족도가 퇴사의 매우 좋은 예측 변수이다. 직업 만족도가 낮을수록 퇴직 확률을 올린다.
  + Argyle, M. (1989). Do happy workers work harder? The effect of job satisfaction on work performance. *How harmful is happiness*, 94-105.

**7)\*딥러닝 모델의 불확실성이 실제로 유익한 데이터 인가(AL시나리오)?**

- uncertainty 기반의 Active learning을 진행헀을 때, 성능이 향상된다.

=> 귀납적으로 이해를 해야하는 느낌.

**<DL 모델의 장점>**

Due to the huge variations of the diseases, it may be sometimes difficult for the humans to analyze the images appropriately. Moreover, inherent limitations of the humans prevent the experts to explore the hidden patterns from the biomedical images. To reduce the human effort and the inherent errors, manual investigations can be replaced to some extent with the computer assisted diagnosis.

To reduce the human effort and the inherent errors, manual investigations can be replaced to some extent with the computer assisted diagnosis

Machine learning methods are used in different stages of image analysis to reduce the human intervention as well as to improve the quality of the results.

* Chakraborty, S., & Mali, K. (2023). An overview of biomedical image analysis from the deep learning perspective. *Research Anthology on Improving Medical Imaging Techniques for Analysis and Intervention*, 43-59.
* DL 서포트 시스템을 도입하는 이유. 업무 과중 외에 업무 성과 측면 부각
  + 1) 숨겨진 패턴을 찾게 해준다
  + 2) 업무량 감소 (임상의 개입 감소)
  + 3) Human error을 감소
  + 4) 진단 결과 정확도 향상

 ML and AI facilitate and assist doctors in diagnosing and predicting the risk of diseases accurately and more rapidly, allowing them to be detected earlier. These techniques enhance the abilities of doctors and researchers to understand how to analyze the generic variations which will lead to disease.

Today, several deep learning-based computer vision applications are performing even better than humans; that is, they are able to identify indicators for cancer in blood and tumors in MRI scans.

* Razzak, M. I., Naz, S., & Zaib, A. (2018). Deep learning for medical image processing: Overview, challenges and the future. *Classification in BioApps*, 323-350.
* ML / AI의 장점
  + 1) 보다 빠르게
  + 2) 보다 정확하게/ 일부 ML은 이미 사람보다 정확도가 높아졌다.
  + 3) 조기에 식별 가능하다.

Machine learning methods coupled with the image processing techniques make the computer assisted diagnosis more powerful and reliable

* Greenspan, H., Van Ginneken, B., & Summers, R. M. (2016). Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE transactions on medical imaging*, *35*(5), 1153-1159.
* 슬슬 신뢰도와 성능을 갖춰가고 있다.

In the current clinical practice, accuracy of detection and diagnosis of cancers and/or many other diseases depends on the expertise of individual clinicians (e.g., radiologists, pathologists) ([Kruger et al., 1972](https://www.sciencedirect.com/science/article/pii/S1361841522000913" \l "bib0002)), which results in large inter-reader variability in reading and interpreting medical images.

Computer-aided quantitative image feature analysis can help overcome many negative factors in clinical practice, including the wide variations in expertise of the clinicians, potential fatigue of human experts, and lack of sufficient medical resources.

* Chen, X., Wang, X., Zhang, K., Fung, K. M., Thai, T. C., Moore, K., ... & Qiu, Y. (2022). Recent advances and clinical applications of deep learning in medical image analysis. *Medical Image Analysis*, 102444.
* 사람이 진단했을 때의 문제점 : 각 임상의의 전문성에 따라 결과 컬리티가 달라진다. [가변성이 크다]
  + DL 서포드 시스템을 통해서
  + 1) 임상의의 전문 지식의 광범위한 차이,
  + 2) 병리학자의 잠재적 피로,
  + 3) 충분한 의료 자원의 부족 등
  + 임상 실습의 많은 부정적인 요소를 극복하는 데 도움이 될 수 있다.

**<Human error 방지 간(Quality Check) 간 DL 시스템 장점>**

**the lack suitable datasets** is one of the biggest barriers to the success of deep learning in medical imaging. On the other hand, the development of large medical imaging data is quite challenging, as annotation requires extensive input from medical experts; in particular, multiple expert opinions are required to overcome the problem of human error.

* Razzak, M. I., Naz, S., & Zaib, A. (2018). Deep learning for medical image processing: Overview, challenges and the future. *Classification in BioApps*, 323-350.
* 기존에 Human error을 막기 위해선 다수의 전문가 의견을 필요로 했다. QC 딴에서 DL 서포트 시스템이 효과가 있다.

We ran a simulation in which the AI system participated in the double-reading process that is used in the UK, and found that the AI system maintained non-inferior performance and reduced the workload of the second reader by 88%. This robust assessment of the AI system paves the way for clinical trials to improve the accuracy and efficiency of breast cancer screening.

Through simulation, we suggest how the system could obviate the need for double reading in 88% of UK screening cases, while maintaining a similar level of accuracy to the standard protocol. These analyses highlight the potential of this technology to deliver screening results in a sustainable manner despite workforce shortages in countries

* McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., ... & Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*(7788), 89-94.
  + => Second leader의 업무량을 88% 가량 줄어준다. Double reading이란 QC와 동일
  + 이는 국가의 인력 부족에도 불구하고 지속 가능한 방식으로 심사 결과를 제공할 수 있는 잠재력을 보여주는 것이다.

**<DL 모델의 한계점 in medical field>**

### **Black-Boxes and Their Acceptance by Health Professionals**

Health professionals are wary as many questions are still unanswered and deep learning theories have not yet provided a complete solution. Even though the deep learning-based method has achieved great success, a solid theory regarding deep learning algorithms is still absent.

most of the researchers using it without knowing the working process or why it provides better results

* Chakraborty, S., & Mali, K. (2023). An overview of biomedical image analysis from the deep learning perspective. *Research Anthology on Improving Medical Imaging Techniques for Analysis and Intervention*, 43-59.
* DL 모델에 대해서 이론적인 뒷받침이 없어 신뢰하지 않는다.

Ultimately, clinicians care about whether the use of algorithms would bring about a beneficial change in patient care, rather than the performance gains reported in papers ([Kelly et al., 2019](https://www.sciencedirect.com/science/article/pii/S1361841522000913" \l "bib0308)).

* Chen, X., Wang, X., Zhang, K., Fung, K. M., Thai, T. C., Moore, K., ... & Qiu, Y. (2022). Recent advances and clinical applications of deep learning in medical image analysis. *Medical Image Analysis*, 102444.
* 임상의들은 실제로 환자들을 케어하는 데 도움이 되는 것을 원한다.

Deep learning, despite being intensively used for analyzing medical images in academia and industrial research institutions, has not made that significant impact as we expected in clinical practice.

* Chakraborty, S., & Mali, K. (2023). An overview of biomedical image analysis from the deep learning perspective. *Research Anthology on Improving Medical Imaging Techniques for Analysis and Intervention*, 43-59.
* 기대한 만큼 Clinical practice 에서 DL 모델의 성능이 잘 안 나오고 있다.

 For the **classification** task, the key to success lies in extracting highly [discriminative features](https://www.sciencedirect.com/topics/computer-science/discriminative-feature) with respect to certain classes. This is relatively easy for domains with large inter-class variance (e.g., accuracies on many public chest X-ray datasets often exceed 90%), but it can be difficult for domains with high inter-class similarity.

For example, the performance of mammogram classification is not so good overall (e.g., 70∼80% accuracies are commonly seen on private datasets), since discriminative features for breast tumors are difficult to capture in the presence of overlapping, heterogeneous [fibroglandular tissues](https://www.sciencedirect.com/topics/engineering/fibroglandular-tissue" \o "Learn more about fibroglandular tissues from ScienceDirect's AI-generated Topic Pages) ([Geras et al., 2019](https://www.sciencedirect.com/science/article/pii/S1361841522000913" \l "bib0279)).

* Chen, X., Wang, X., Zhang, K., Fung, K. M., Thai, T. C., Moore, K., ... & Qiu, Y. (2022). Recent advances and clinical applications of deep learning in medical image analysis. *Medical Image Analysis*, 102444.
* 유사한 조직 구조를 띈 경우 학습이 어렵다.
* 유방암 같은 구조에 대해서는 정확도가 70~80% 밖에 나오지 않는다.

 The notion of fine-grained visual classification (FGVC) ([Yang et al., 2018](https://www.sciencedirect.com/science/article/pii/S1361841522000913" \l "bib0280)), which aims at identifying subtle differences between visually similar objects, might be suited for learning distinctive features given high inter-class similarity.

* 유사한 것 사이의 미묘한 차이를 포착하도록 만들 수 있다.
* 단, FVGC 데이터셋은 클래스 간 높은 유사성을 띄도록 의도적으로 수집되었음. 결과적으로, 이러한 데이터 세트에 대해 개발되고 평가된 접근 방식은 모든 이미지가 아닌 특정 부분만 높은 클래스 간 유사성을 나타내는 의료 데이터 세트에 쉽게 적용되지 않을 수 있다.

However, after systematically reviewing over 200 prediction models from 169 studies that were published up to 1 July 2020, [Wynants et al. (2020)](https://www.sciencedirect.com/science/article/pii/S1361841522000913" \l "bib0302) concluded that all these models were of high or unclear risk of bias, and thus none of them were suitable for clinical use – either moderate or excellent performance was reported by each model; however, the optimistic results were highly biased due to model overfitting, inappropriate evaluation, use of improper data sources, etc.

Similar conclusion was drawn in another review paper ([Roberts et al., 2021](https://www.sciencedirect.com/science/article/pii/S1361841522000913" \l "bib0303)) – after reviewing 62 studies that were selected from 415 studies the authors concluded that, because of methodological flaws and/or underlying biases, none of the deep learning and machine learning models identified were clinically applicable to the diagnosis/prognosis of COVID-19

* Chakraborty, S., & Mali, K. (2023). An overview of biomedical image analysis from the deep learning perspective. *Research Anthology on Improving Medical Imaging Techniques for Analysis and Intervention*, 43-59.
* 현재 DL 서포트 시스템은 임상 실험에 신뢰받지 못하고 있다고 판단된다.
* 과적합, 부적절한 데이터 소스, 등 다양한 이유로 모델 사용 부적합 판단을 받았다.

Going beyond the example of COVID-19, the high-risk bias of deep learning approaches is indeed a recurring concern across different medical image analysis tasks and applications ([Nagendran et al., 2020](https://www.sciencedirect.com/science/article/pii/S1361841522000913" \l "bib0304)), which has severely limited deep learning's potential in [clinical radiology](https://www.sciencedirect.com/topics/medicine-and-dentistry/clinical-radiology). Although quantifying the underlying bias is difficult, it can be reduced if handled appropriately.

* Chakraborty, S., & Mali, K. (2023). An overview of biomedical image analysis from the deep learning perspective. *Research Anthology on Improving Medical Imaging Techniques for Analysis and Intervention*, 43-59.
* 편향이 모델을 신뢰할 수 없는 주요 이유이며, 이를 조절한다면 충분히 성능을 향상시킬 수 있다.

we need to be cautious about the potential biases caused by using a single public dataset alone – as the whole community strive for achieving state of the art performance, community-wide overfitting is likely to exist on this dataset ([Roberts et al., 2021](https://www.sciencedirect.com/science/article/pii/S1361841522000913#bib0303)).

* 단일 데이터셋에서만 학습했을 때, 과적합으로 인한 잠재적 편향에 대해 주의를 기울여야 한다.

Despite the success, the further improvement of deep learning models in medical image analysis is majorly bottlenecked by the lack of large-sized and well-annotated datasets. In the past five years, many studies have focused on addressing this challenge.

the further performance improvement is majorly hurdled by the requirement for large amounts of annotated datasets.

As compared to regular datasets in [computer vision](https://www.sciencedirect.com/topics/engineering/computervision), a medical image dataset usually contains relatively small amounts of images (e.g., less than 10,000), and in many cases, only a small percentage of images are annotated by experts.

* Chen, X., Wang, X., Zhang, K., Fung, K. M., Thai, T. C., Moore, K., ... & Qiu, Y. (2022). Recent advances and clinical applications of deep learning in medical image analysis. *Medical Image Analysis*, 102444.
* 가장 큰 문제가 데이터셋을 확보하는 문제다.
* 데이터셋이 충분하지 않을 때 성능이 크게 악화한다.
  + 필요로 하는 데이터셋의 크기를 제시해준다면, AL이 얼마나 데이터 양을 줄여주는 지 가시화할 수 있지 않을까?
* 데이터가 10000개 보다 적은 사례가 많다. => 최소 1만개는 되어야 학습하기에 충분하다?
  + + 의사들이 주석을 단 경우는 그보다 더욱 작다.
  + 컴퓨터 비전 쪽의 데이터셋과 비교를 해주면 일반적으로 얼마나 필요로 한지 보여줄 수 있을까?