

画像診断支援 AI を持続的に開発・性能評価する 基盤環境の構築に関する多施設共同観察研究

1. 研究の概要と対象

本研究は、国立がん研究センター研究倫理審査委員会で承認され、各参加機関の長の許可を受けて実施される多機関共同研究です。2000 年 4 月 1 日～2026 年 3 月 31 日までの間に、国立がん研究センター中央病院、がん研有明病院、神奈川県立がんセンター、静岡県立静岡がんセンター、愛知県がんセンター、兵庫県立がんセンターにおいて、様々な臓器にがんを有する患者のうち、生体試料以外の医療情報である放射線画像検査(CT、MRI、PET-CT/MRI、及び X 線写真等)を受けた 18 歳以上の患者さんの放射線画像データを収集します。収集した放射線画像データを用いて、施設間で生じるデータの偏りに対して性能が低下しないことが保証された、人工知能技術の開発を目指します。

対象者:胸部腫瘍(特に前縦隔腫瘍)

対象期間:2000 年 04 月 01 日～2026 年 03 月 31 日

予定:最大 3000 例

2. 研究目的・方法

研究実施期間:研究許可日から 2029 年 3 月 31 日まで。

人工知能技術の活用は多くの分野で急速に進行しており、医療も例外ではありません。特に、CT や MRI などの医用画像から疾患の診断を行うプロセスは、人工知能技術との親和性が高く、一部の疾患では既に実用化もされています。このような人工知能技術を用いた臨床支援システムは、高度な専門性のために人的リソースが不足し、大きな負担を抱えている臨床現場における解決策の一つになるのではないかと期待されており、世界中で活発な研究開発が展開されています。

しかしながら、現状の人工知能技術を安全に臨床現場で用いるためには、まだまだ解決すべき根本的な課題が山積しています。そのうちの 하나가、人工知能技術を用いて作ったモデル(ソフトウェア)が、特定のデータの集合に偏って学習してしまう過学習と呼ばれるものです。こうした過学習は、例えば、施設ごとに異なる疾患の出現頻度や、医用画像の撮像機器や撮像方法の違いなどによっても生じます。仮に過学習が生じてしまうと、当該モデルを訓練する際に与えたデータの集合に属さない新規のデータに対する性能が大きく低下することが知られています。そのため、例えば、ある病院のデータを使って訓練したモデルを他の病院のデータに当てはめた際に、疾患の検出性能が大きく低下してしまい、その予測結果を信頼することができなくなるといった問題が発

生じます。このような過学習は、特に大量のデータを収集することが困難な医療分野で起こりやすく、人工知能技術を普及させるに当たっての信頼性・安全性に関わる大きな問題になっています。

本研究では、国立がん研究センター中央病院・研究所と、がん研有明病院、神奈川県立がんセンター、静岡県立静岡がんセンター、愛知県がんセンター、兵庫県立がんセンターの共同研究開発体制を構築し、このような学術的課題に取り組みます。具体的には、施設間で生じうるデータの偏りに対して性能が低下しないことが保証された、人工知能技術の開発を目指します。本研究では、症例が集約できず十分に画像診断支援 AI 開発が実施できない希少がんを中心に、広くがんを対象にした画像診断支援 AI 開発を行います。こうした学術的知見は、将来的に、様々な施設で利用可能な人工知能技術の開発に繋がり、未来の患者さんに対して、がんの早期の発見と治療を支援できる可能性があります。研究方法は、これらの情報を人工知能技術の一つである機械学習・深層学習 (Deep Learning) を用いて解析します。

3. 研究に用いる試料・情報の種類

本研究で用いる診療情報は、DICOM データと呼ばれる放射線画像の情報が含まれたデータと、これに付随する臨床情報として、性別、原発腫瘍の診断及び治療に関わる情報 (診断日、診断時年齢、疾患名、ステージング、治療内容等)、既往症、合併症、がんの診断及び治療に関わる情報 (診断日、診断時年齢、全身状態、がんに対する治療内容等) 等になります。各施設あたり、胸部腫瘍 (特に前縦隔腫瘍) と診断された放射線画像を最大 3,000 症例まで収集します。

【国立がん研究センター中央病院の患者さん】

DICOM データは、個人を特定しうる情報として、診療録 ID、氏名、生年月日、病院名が含まれます。今回の研究では、診療録 ID、氏名、は削除して使用しません。尚、個人識別情報である生年月日は、国立がん研究センターに研究用に敷設されたネットワーク内で、使用します。このネットワークは、外部のインターネットから一切遮断され、病院カルテと同じレベルのセキュリティが担保されているため、個人識別情報が漏洩する心配はありません。

【がん研有明病院、神奈川県立がんセンター、静岡県立静岡がんセンター、愛知県がんセンター、兵庫県立がんセンター、の患者さん】

DICOM データは、個人を特定しうる情報として、診療録 ID、氏名、生年月日、病院名が含まれます。今回の研究では、診療録 ID、氏名、生年月日、病院名は削除して使用しません。

本研究では、生体試料は用いません。上記の手続きにより匿名化した DICOM データについては、外部のネットワークから隔離された専用端末で保管、解析します。これにより、データの漏洩のリスクを最小化するとともに、万が一の漏洩時にもデータそれ自体から特定の個人を識別することができないようにします。

4. 外部への情報の提供

研究全体で収集したデータは国立がん研究センター研究所において、セキュリティの担保されたパソコン等にデータを保管し、厳重に管理します。がん研有明病院、神奈川県立がんセンター、静岡県立静岡がんセンター、愛知県がんセンター、兵庫県立がんセンターから国立がん研究センター研究所へのデータの提供は、データを匿名化した上で暗号化し、特定の研究者以外がアクセスできない記録媒体を郵送等により受け渡すことを行います。

匿名化されたデータと元の患者さんの個人情報を紐付けるための対応表は、データ提供元の研究責任者が保管・管理します。

5. 将来の研究における情報の二次利用

本研究で用いた試料・情報を国内外の機関で実施する将来の医学的研究のために、研究終了後も大切に保管させていただきます。新たな研究に用いる際には、国内外の規制に則り、あらかじめ研究計画書を作成して研究倫理審査委員会の承認や研究機関の長の許可を受ける等、適正な手続を踏んだ上で行います。なお、新たな研究の概要・研究機関については、新たな研究に関わる機関(試料・情報の授受を行う機関すべて)公式ホームページ等にて情報公開いたします。

国立がん研究センターが参加する研究の公開情報については国立がん研究センターの公式ホームページより確認することができます。<https://www.ncc.go.jp/jp/>

6. 研究組織

研究代表者

国立がん研究センター 中央病院放射線診断科:渡辺裕一

研究責任者

国立がん研究センター 中央病院放射線診断科:渡辺裕一

がん研有明病院 画像診断センター:寺内隆司

静岡県立静岡がんセンター 画像診断科:遠藤正浩

兵庫県立がんセンター 放射線診断 IVR 科:竹中大祐

神奈川県立がんセンター 放射線診断 IVR 科:栗原宏明

愛知県がんセンター 放射線診断・IVR 部:佐藤洋造

7. お問い合わせ先

本研究に関するご質問等がありましたら下記の連絡先までお問い合わせ下さい。

ご希望があれば、他の研究対象者の個人情報及び知的財産の保護に支障がない範囲内で、研究計画書及び関連資料を閲覧することが出来ますのでお申出下さい。

また、試料・情報が当該研究に用いられることについて患者さんもしくは患者さんの代理人の方にご了承いただけない場合には研究対象としますので、下記の連絡先までお申出ください。その場合でも患者さんに不利益が生じることはありません。

照会先および研究への利用を拒否する場合の連絡先:

〒135-8550 東京都江東区有明 3-8-31

Tel:03-3520-0111 (代表)

がん研有明病院 画像診断センター: 寺内隆司