

II. 分担研究報告

(1) 研究分担者報告書

病理診断学・サルコーマに関する研究

研究分担者 高澤 豊 がん研究会がん研究所 病理部副部長

研究要旨:「病理デジタル画像データの深層学習・人工知能による病理画像認識診断支援ツールの開発研究」では稀少がんの診断支援が重要なテーマである。がん研究会有明病院では、稀少がんのうち肉腫症例が多く、これらの症例を活用することを目的として、データベースの作成、デジタル画像ファイルの蓄積を試みた。

がん研有明病院に蓄積された約 23,000 例の骨および軟部腫瘍を WHO 分類 (2013) に基づいてデータベースを作成した。WHO 分類の大項目の 125 組織型すべてについて、症例があることが確認された。各組織型の亜型を含め、約 200 の組織型に分類した。各組織型の代表例を抽出し、組織像の確認したのち、典型的な部分の組織プレパラートをスキャナーで読み込み、デジタル画像を取得し、ファイルサーバ上に保存した(500 症例)。

WHO 分類と過去の組織分類が必ずしも WHO 分類に対応しておらず、分類不能な症例も相当数含まれていた。いずれも稀少がんとされる肉腫であり、それらを AI でどのように扱うべきか、どう学習させるのか、ということが来年度の課題である。また、分類不能な稀少がんについては、症例をより蓄積することによって、新しい疾患概念が生まれる可能性もあり、腫瘍病理学の基礎研究としても意義がある。

A. 研究目的

「病理デジタル画像データの深層学習・人工知能(以下 AI)による病理画像認識診断支援ツールの開発研究」では、術中迅速病理診断と稀少がんの診断支援が最も重要なテーマである。稀少がんは病理診断自体が難しいもの、言い換えると、診断者間の不一致が少なくないという問題点が存在する。がん研究会有明病院では、稀少がんのうちとりわけ肉腫症例が多く集まっており、これらの症例を活用することによって「AI 活用による病理診断支援ツール開発」が期待される。

B. 研究方法

1. 稀少がんの症例選択

がん研有明病院に蓄積された約 23,000 例の骨および軟部腫瘍を WHO 分類 (2013) に基づいて再分類し、稀少がんを抽出を行う。必要に応じて、免疫組織化学的検索や遺伝子検索を追加し、診断確認する。

2. 稀少がんの病理組織のデジタル画像化 DB から各疾患の代表例を抽出し、組織像の確認したのち、各症例ごとに最も典型的な部分の組織プレパラートを選択し、スキャナーで読み込み、デジタル画像をサーバ上に保存する。

(倫理面への配慮)データベースはがん研有明病院に蓄積された全ての症例のものであり、院内のみで閲覧可能な診療システム上に作成している。デジタル画像としてアーカイブ化する症例は検体の研究、

教育使用に関する患者本人の承諾を得られている症例であり、デジタル画像については個人情報を含まない。遺伝子検索は腫瘍組織の体細胞変異に限って行い、通常の診断過程で検索されるものを解析する。平成29年度は希少がんの画像データの院外での解析に当たっては、がん研の臨床研究委員会での審査、承認を得たのちに研究に用いる予定である。

C．研究成果

1．希少がんの症例選択

がん研における骨軟部腫瘍データベース(DB)の更新＝約23,000例の再整理、更新を行った。希少がんの多い骨軟部腫瘍は近年蓄積されている分子病理学的知見により、分類、疾患名が変化しており、ICD10、SNOMED-CTなどでは多種多様な骨軟部の腫瘍特に希少がんを1対1対応でコード付けできないので、データベースを作成する際の基礎となる組織学的分類は、現在最も標準的な分類であるWHO分類(2013)を用いた。その結果WHO分類の大項目の125組織型すべてについて、症例があることが確認された。各組織型の中の亜型を含めると、約200の組織型に分類できた。データベースに含まれる情報は、組織型(および亜型)、年齢、性別、既往歴、家族歴、腫瘍の部位、腫瘍の大きさ、肉眼的特徴、組織像の特徴、腫瘍の免疫組織化学的形質、遺伝子検索の結果である。今年度は、WHO分類の全ての組織型を抽出することに主眼をおいたので、蓄積された約23,000症例のうち約1500症例のデータベースを完成させることができた。

2．希少がんの病理組織のデジタル画像化データベースから各疾患例を抽出し、組織

像の確認ができたものについて、各症例ごとに最も典型的な部分の組織プレパラート(数枚)を選択した。過去の症例で染色の劣化が高度なものについては、標本を再作成した。今年度デジタル画像化したプレパラートは500症例分である。デジタル画像は院内のファイルサーバ上に、個人情報を除いた方で保存した。

D．考案

現在最も標準的な分類であるWHO分類を用いて組織型を決定する過程で、過去の組織分類が必ずしもWHO分類に対応しておらず、分類不能な症例が相当数含まれていた。いずれも希少がんであり、それらをAIでどのように扱うべきか、どう学習させるのか、課題であると考えられた。また、分類不能な希少がんについては、症例をより蓄積することによって、新しい疾患概念が生まれる可能性もあり、実臨床のみならず、腫瘍病理学の基礎研究としても意義あるものである。

E．結論

今年度は、がん研有明病院に蓄積された23000症例の骨軟部腫瘍から、希少がん症例を抽出するためのデータベースを作成した。AIで学習させるためのデジタル画像もアーカイブ化した。組織型決定の困難な希少がんが抽出された。

F．健康危険情報(総括参照)

G．研究発表

1. 論文発表：特になし
2. 学会発表：特になし

H．知的財産権の出願・登録状況

1. 特許取得：特になし
2. 実用新案登録：特になし
3. その他：特になし

(2) 研究分担者報告書

医学統計に関する研究

研究分担者 山口 拓洋 東北大学大学院医学系研究科医学統計学分野・教授

研究要旨：生物統計の観点からバリデーション研究のデザインについて検討した。要配慮個人情報取り扱いについて、個人情報保護法等の規制の整備状況を確認し、希少がんの病理画像の取り扱い方針を検討した。

A．研究目的

ツール開発におけるバリデーション研究デザインの策定、及び、研究デザイン策定の際に必要な個人情報保護等の法令規制について整理する。

B．研究方法

診断性能の評価としてのエンドポイントの設定を中心にサンプルサイズ設計等の研究デザインを検討する。

(倫理面への配慮)

ヘルシンキ宣言、人を対象とする医学系研究に関する倫理指針、及び、改正個人情報保護法に遵守して、実施した。

C．研究結果

・エンドポイントの検討

診断能のエンドポイントとして一般的に用いられる、感度、特異度を主要評価とするが、閾値の設定が問題となる。通常の検定ベースの評価のみならず、ベイズ流の観点からの評価も考慮する。

・サンプルサイズ設計

上述した検定ベースでの閾値と期待値の設定をまずは考慮するが、必要サンプルサイズが現実的に不可能な数字になる可能性が考えられ、こちらについても、ベイズ流の方法なども考慮する。

D．考察

現時点で、バリデーション研究のデザイン策定に必要な情報が十分に取得できていないことから、今後の更なる検討が必要である。

E．結論

生物統計の観点から研究デザインについて検討した。要配慮個人情報の取り扱いについて、個人情報保護法等の規制の整備状況を確認し、希少がんの病理画像の取り扱い方針を検討した。

F．健康危険情報

(分担研究報告書には記入せずに、総括研究報告書にまとめて記入)

G．研究発表

1. 論文発表：特になし
2. 学会発表：特になし

H．知的財産権の出願・登録状況

1. 特許取得：特になし
2. 実用新案登録：特になし
3. その他：特になし

(3) 研究分担者報告書 臨床試験データ管理学・SNOMED-CT に関する研究

研究分担者 宮路天平 東京大学大学院医学系研究科 臨床試験データ管理学講座・

特任助教

研究要旨：SNOMED-CT の詳細調査等を行い、会議の際に情報を提供した。また病理デジタル画像・人工知能技術を用いた、病理画像認識による術中迅速・ダブルチェック・希少がん等病理診断支援ツールの開発における研究デザインの立案を行った。

A．研究目的

SNOMED-CT の詳細調査等を行い本研究での使用に関して助言を行う。また「病理デジタル画像・人工知能技術を用いた、病理画像認識による術中迅速・ダブルチェック・希少がん等病理診断支援ツールの開発」の研究において、臨床データマネジメントの観点から、研究デザインの計画や実施についての研究支援を行う。

B．研究方法

術中の迅速病理診断システムの開発について、コンサルテーションをベースに研究デザインの立案の助言を行った。

また、希少がんの病理診断困難症例の画像認識診断支援ツールの開発については、SNOMED-CT による診断名のマッピングを検討し、SNOMED-CT の階層構造や運用状況の調査を行った。

2017 年 3 月には、改正個人情報保護法に関連して、人を対象とする医学系研究に関する倫理指針が改訂されたため、本研究課題における希少がん患者の症例データの取り扱いについて検討を行った。

(倫理面への配慮)

当研究課題は、ヘルシンキ宣言、人を対象とする医学系研究に関する倫理指針および改正個人情報保護法に遵守して、実施した。

C．研究結果

開発された病理システムについては、妥当性の検証のための Validation 研究を計画することになるため、研究デザイン上考慮すべき事項(エンドポイント、データのサンプリング等)について、助言を行った。

SNOMED-CT の導入については、本邦では十分に活用されておらず、臨床現場への一般化可能性を考慮し導入を見送った。

D．考察

開発中のシステムの仕様や現状の精度に基づき、妥当性および信頼性の検証のための Validation 研究の実施計画書の作成を行う。

E．結論

平成 29 年度に実施する Validation 研究にむけて、研究デザインの立案や助言等、準備を進めることが出来た。

F．健康危険情報

総括研究報告書を参照

G．研究発表

1. 論文発表：なし
2. 学会発表：なし

H．知的財産権の出願・登録状況

1. 特許取得：なし
2. 実用新案登録：なし
3. その他：なし

(4) 研究分担者報告書

深層学習・人工知能に関する研究

研究分担者 野村直之 メタデータ株式会社 代表取締役社長

研究要旨：東大病院より提供されたリンパ節の超高解像度病理画像群を数千から数万枚に分割し腫瘍有無等の正解データを作成。本正解データで深層学習をトレーニングし95%程度の認識精度を達成。

A．研究目的

高解像度病理画像を通常の深層学習に学習させる方式の考案、手順の開発、ツール仕様の策定、小規模実験による方式の有効性の評価。

B．研究方法

各数10億画素の病理画像を数万画素のタイル状画像群数万枚に分割したものに「腫瘍」「健常」「空白」等のラベルを付け深層学習にかけてトレーニング。未知画像を機械的に分割した画像ごとに十分高い判定精度が出ることを確認する。

(倫理面への配慮)

医療機関から受け取る全てのデータについて個人情報情報は削除済。

C．研究結果

8割方人手で延べ100万枚超のタイル画像に切り出した学習用の正解データは、今回購入した36TFlopsミニ・スーパーコンで30時間程度で学習。概ね95%程度の精度が出ていることを確認。

D．考察

今回確認された精度は、29年度早々に開発予定の、タイル画像の元レイアウトで腫瘍他の判定結果の色分け表示(確率値の高いものは赤色等)をするツールにより充

分実用に耐える水準になるといえるのではないかと。特に、タイル1枚のみ赤変しているなど転移領域がごく狭い場合に、人間医師の見落としの低減に貢献できていくものと思われる。今回対象としたリンパ節は、もともと画像としてのバリエーションが少なく、診断も容易で、正解ラベル付けの苦勞が少なく、トレーニング量、学習時間も概ね妥当なものに収まった。今後は、多彩な絵、多種への分類に備えて、正解データ作りの自動化率を高めたり分割ファイルの扱い、計算パワーの効率活用の工夫が必要となるだろう。

E．結論

高解像度病理画像を深層学習で扱う今回の方式は工夫次第で様々な癌の診断支援に効果的に活用可能になると思われる

G．研究発表

1. 論文発表：特になし
2. 学会発表：特になし

H．知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得：(取得の予定・発表を急いで他機関の後願を排除の予定)
2. 実用新案登録：特になし
3. その他：特になし

(5) 分担研究報告書

病理デジタル画像加工技術に関する研究

研究分担者 宮越 徹 インスペック株式会社 開発部 社員

研究要旨：本研究では、人工知能が学習可能な画像を、バーチャルスライド装置による病理デジタル画像から作成可能なことが分かった。具体的には、学習に必要な画像の解像度や形式で必要な画像領域を抽出保存することで、目的の画像が得られることが分かった。

A．研究目的

人工知能技術を用いた病理画像認識による術中迅速・ダブルチェック・希少がん等病理診断支援ツールの開発において、人工知能が学習可能な病理画像を作成することを本研究の目的とする。

B．研究方法

人工知能が学習するための画像の解像度や形式を理解し、バーチャルスライド装置により高解像度(1.5cm 四方、対物 20 倍レンズで 9 億画素、対物 40 倍レンズで 36 億画素)でデジタル化された病理標本から、適切な画像を切出すツールを開発することにより本研究を行う。

(倫理面への配慮)

本研究に使用する病理標本は連結可能な匿名化により個人情報情報が反映されないよう配慮している。

C．研究結果

デジタル病理標本の表示画面上で、操作者が指定する任意の正方領域を下記の解像度及び記録形式のいずれかにて保存するツールを開発し、人工知能が学習するための画像作成が可能となった。

解像度(画素)：64x64,128x128,
256x256,512x512

記録形式：BMP,JPG,PMG,TIFF

D．考察

人工知能の学習に適切な解像度及び記録形式で画像を切出して保存することで、高解像度でデジタル化された病理標本から、人工知能の学習用の画像を作成可能であると考えられる。

E．結論

本研究により、人工知能が学習するための画像を、デジタル病理標本から作成可能であることが分かった。人工知能の学習には出来るだけ多くの画像が必要なため、今後は、学習用画像を自動で複数枚作成するツール等の開発を検討し、さらなる高効率化を図りたい。

G．研究発表

1. 論文発表：特になし
2. 学会発表：特になし

H．知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得：特になし
2. 実用新案登録：特になし
3. その他：特になし