

AIによる画像診断支援に向けた 研究の進捗状況について

国立研究開発法人

日本医療研究開発機構（AMED）

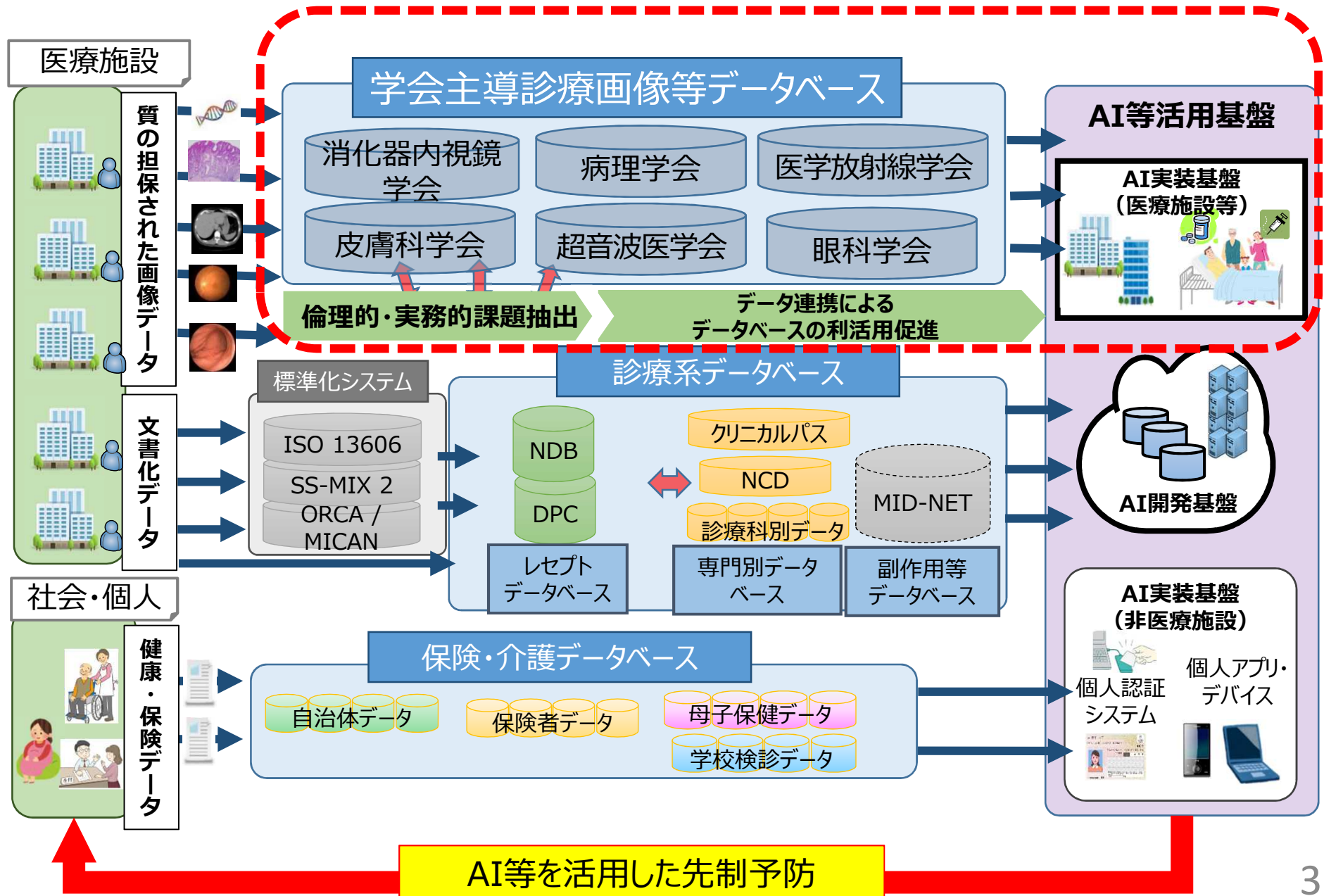
理事長 末松 誠

平成30年7月23日
第1回保健医療分野AI開発加速コンソーシアム
於：厚生労働省

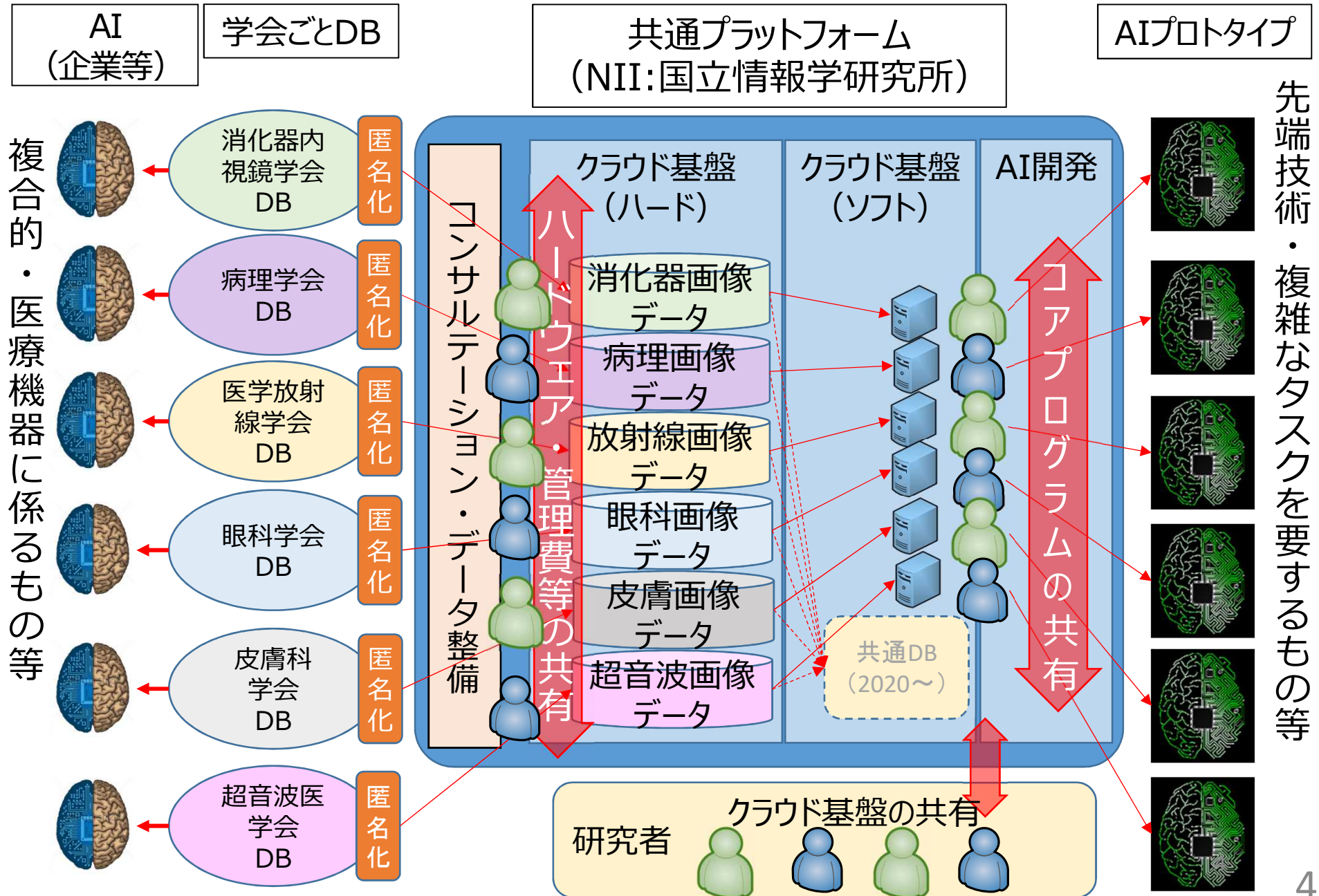
AMEDにおける「骨太の取り組み」

1. 認知症の包括的対策（その内、研究開発推進に関わるもの）
2. 子どもの健全な成育と疾患克服に資する研究
3. AI等を活用した先制医療・予防医療への取り組み
4. AROネットワーク機能の強化
5. 遺伝子治療・ゲノム編集等の基礎臨床研究基盤整備
6. 感染症及びがん研究での新規シーズ創出（若手研究者育成）
7. 感染症領域でのグローバルヘルスに資する研究開発

AI等を活用した先制医療・予防医療への取り組み



臨床画像情報基盤の全体像



共通プラットフォーム（NII：国立情報学研究所）の役割

◆コンサルテーション

- 各学会がAIにより実現したい内容の整理
- AI対象疾患の定義と診断の範囲の決定
- 対象画像の規格の整理
- 画像に付帯すべき関連情報（メタデータ）の種類と粒度の決定

◆AIを用いた画像解析に必要なクラウド基盤の整備

- クラウドハード面の整備
- クラウドソフト面の整備：メタデータの格納様式の共通化
- 各学会のクラウド使用マニュアルの作成

◆AIプロトタイプ構築

- 収集されたデータを用いたAIプロトタイプ開発
- 学会によるAIプロトタイプの試用とフィードバックによる改良

コンサルテーションの例：データのアノテーションについて

- ◆現在の診療画像AI開発に、アノテーション※は必須である。
- ◆開発するAIごとに必要とするメタデータの種類と粒度が異なる。

- 異常・正常のみ
- 診断名
- 組織名 画像では特に重要

- 病変部位のマーキング
 - 矩形による囲い込み
 - 細胞単位の囲い込み

見落とし防止



より複雑な診断補助

- ◆共通に使用されるメタデータは同一形式で格納する必要がある。
- ◆専門家の中でアノテーションに対する認識を共有する必要がある。

※アノテーション：AI開発に必要な画像関連情報（メタデータ）を付与すること

メタデータの粒度の違い（例）

◆病理学会:不均一かつ複雑な画像

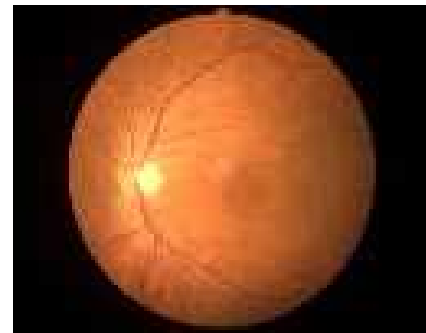
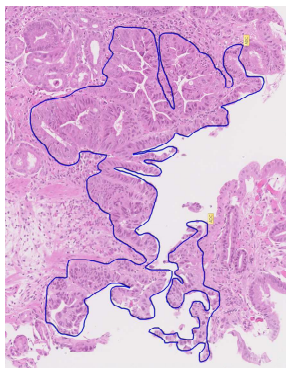
- 病変につき、細胞単位の細かいマーキングが必要
- 少数の標本で比較的精度の高いAIの開発が可能

◆消化器内視鏡学会:比較的均一な画像

- 病変部は四角など粗いマーキングで可
- 多数のデータを学習させることで精度を向上

◆眼科学会:非常に均一な画像

- マーキングは行わず、病名と眼底写真のみで学習（緑内障）
- 今後全身疾患（糖尿病・認知症など）の診断支援AIのためには、詳細なテキストデータが必要



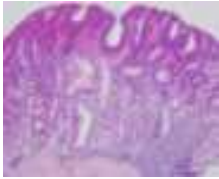
臨床画像情報基盤の採択状況

	H28	H29	H30	H31	H32
病理学会	← H28補正 →		← H30本予算～ →		
消化器内視鏡学会	← →		← →		
医学放射線学会	← →		← →		
眼科学会	← H29調整費～ →				
超音波医学会			← H29補正 →		
皮膚科学会			← →		
国立情報学研究所			← →		

平成29年度までの成果例

1. データベース基盤整備
2. AI開発
3. 学会間連携
4. 共通プラットフォーム構築
5. 学会共通課題の解決への取り組み

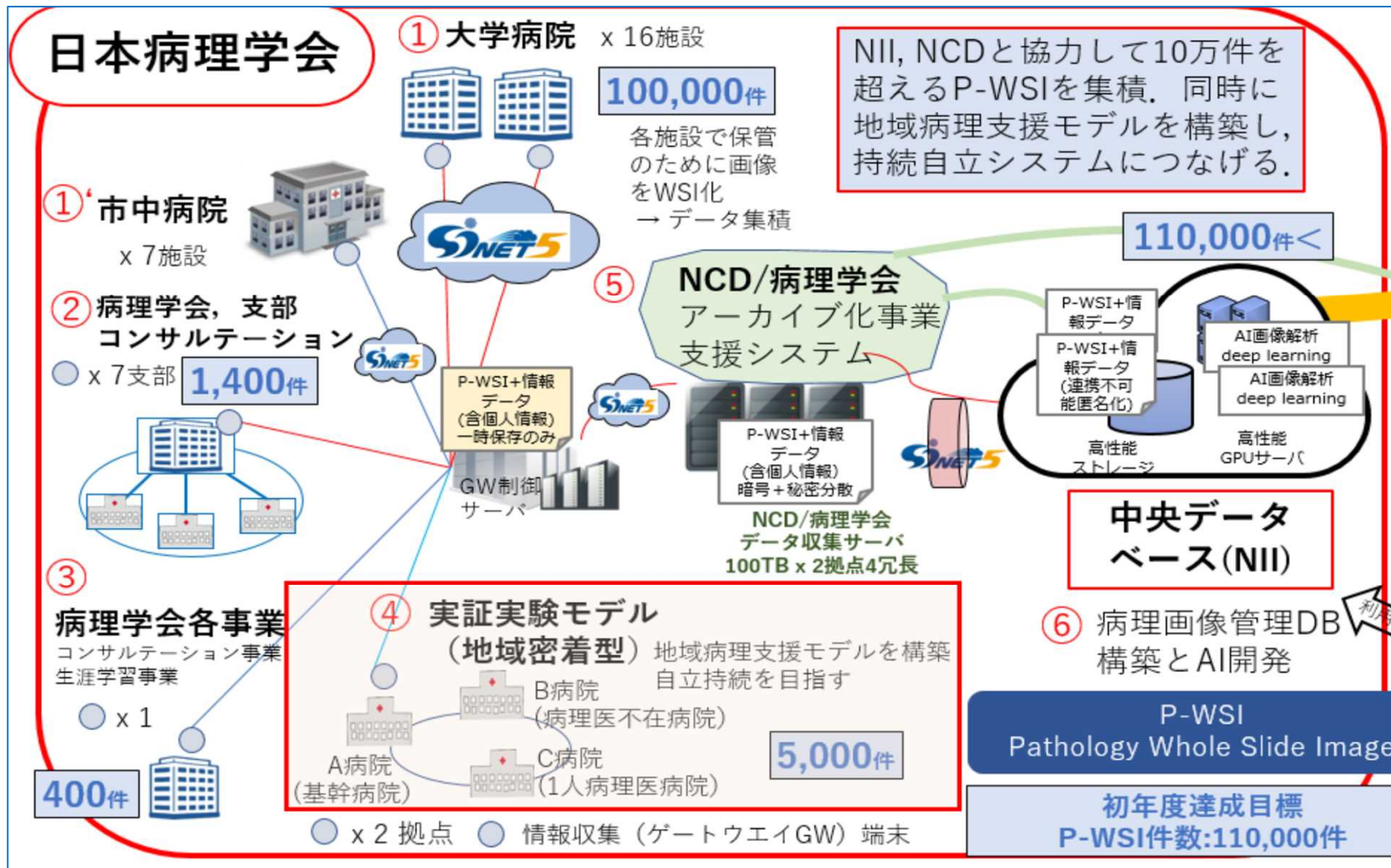
事後評価委員会(平成30年3月13日開催)、事前評価委員会(5月8日)資料を基に作成

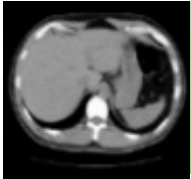


平成29年度までの成果 データベース基盤整備①

病理

- ◆データ登録用クラウドサーバー・基幹回線の整備
- ◆日本でP-WSIスキャナーを販売している8社の共通ビューを開発



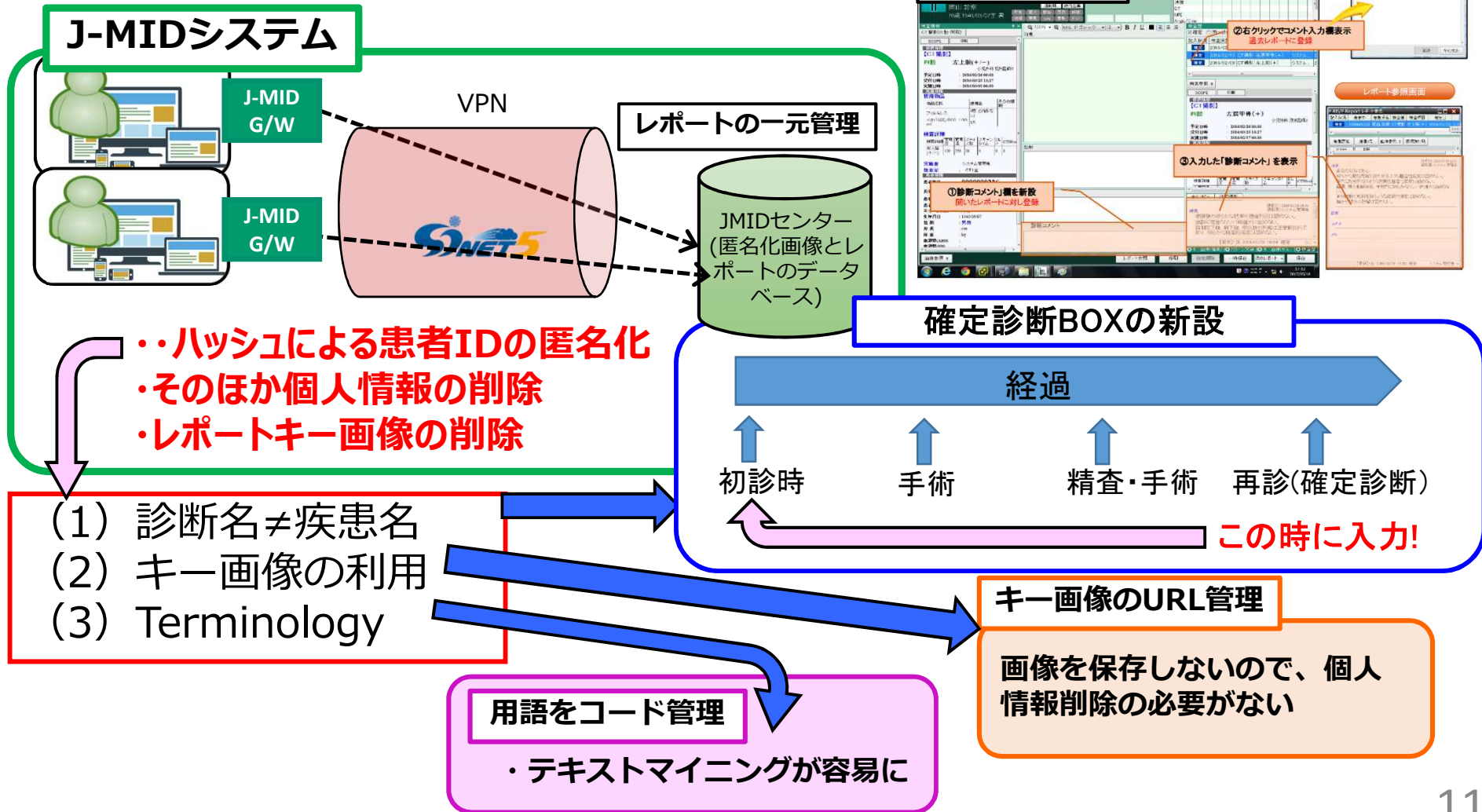


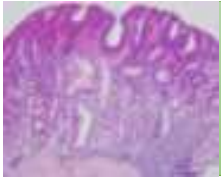
平成29年度までの成果 データベース基盤整備②

放射線

◆ 画像の一元管理によるAI診断の教師データの作成補助

- 画像とレポートの紐付け
- レポート文章内の単語をコード管理



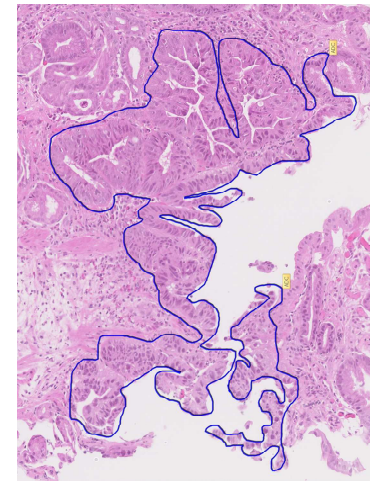
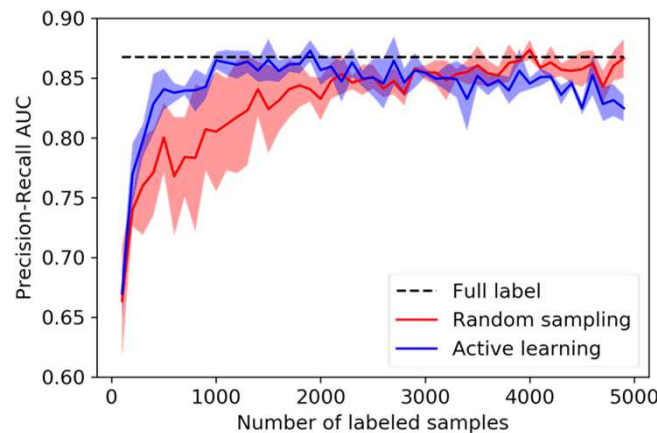


平成29年度までの成果 AI開発①

NII

病理

- ◆ 胃がん標本を用い正常／異常識別を行うパイロットシステム
 - 20症例（5,000枚の切出し画像）の WSI ※¹を用いた学習・評価にてAUC 0.87 を達成※²



能動学習による少量学習データからの効果的学習

※開発の時点でアノテーション作業がボトルネックとなることが判明
従来のアノテーションにかかる時間：1標本当たり1～3時間
→スキャナーメーカーを問わない作業効率の高いツールを開発

※¹ WSI : Whole Slide Image

※² 7月13日現在は80症例に増加、AUCは0.912に改善



平成29年度までの成果
AI開発②

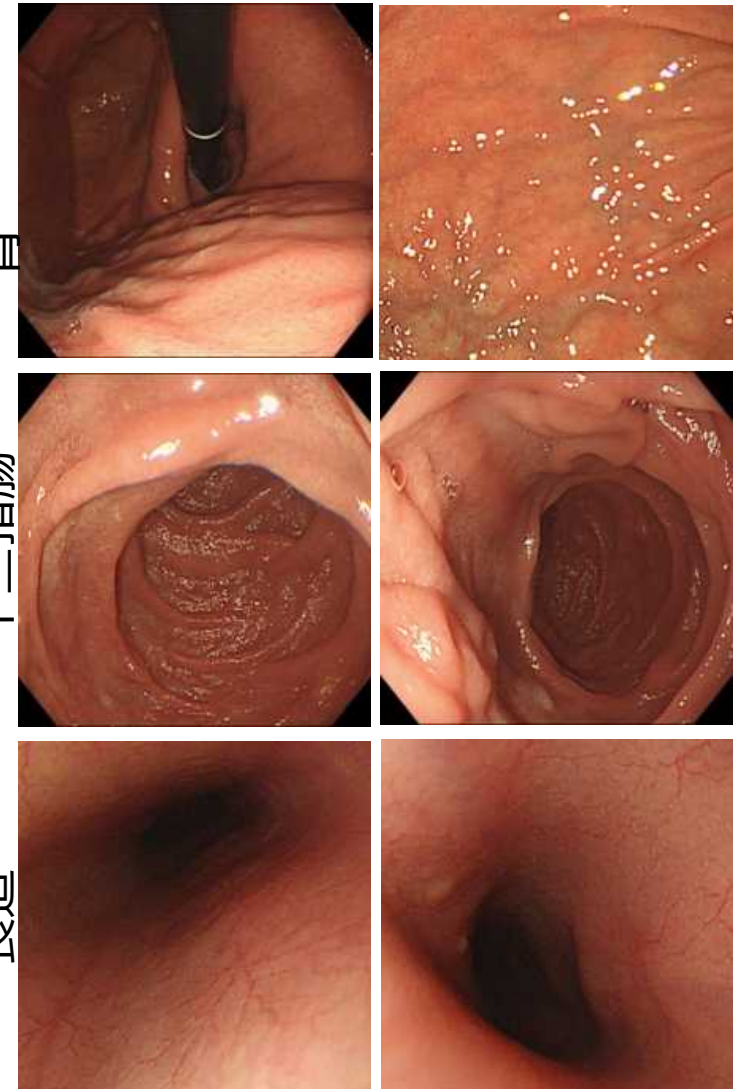
NII

内視鏡

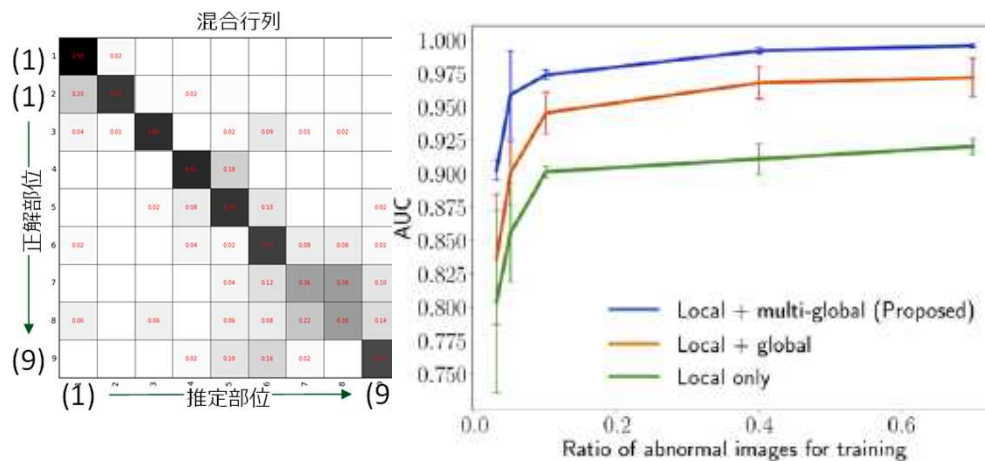
◆ 胃内視鏡画像の臓器自動分類

102,006画像を分類し、平均99.8%
の分類精度を達成

- 胃 99.9%
- 十二指腸 99.8%
- 食道 99.8%



分類結果





平成29年度までの成果 AI開発③

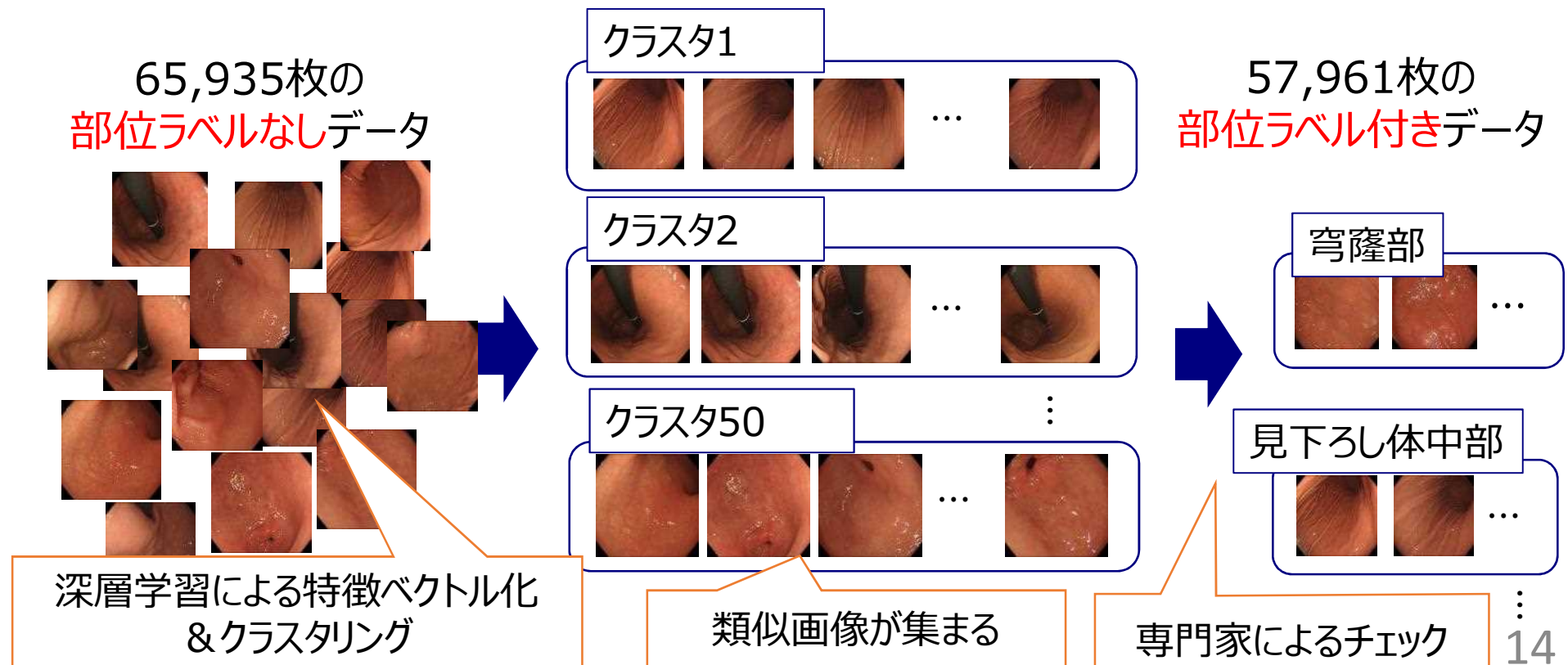
NII

内視鏡

◆上部消化管内視鏡の自動部位特定システム

- 教師なしデータによる自動画像クラスタリングにより、50クラスタへ分類
- 医師分類 (20クラス) を加え、9クラスへ統合
→80%の認識精度を達成

今年度より企業と連携して製品開発予定



平成29年度までの成果 学会間連携

病理

内視鏡

放射線

- ◆学会共通ハッシュ化IDによる連携
- ◆学会共通ビューワの構築

マッチング実行

マッチング処理を実行します。
この処理は時間がかかる場合があります。

実行 閉じる

データファイル C:¥Temp¥病理データ¥201712¥病理テストデータ20...

検査日フィールド名 採取日

患者IDフィールド名 ID

静止画フォルダ C:¥Temp¥病理データ¥201712

DICOMフォルダ C:¥Temp¥DICOM¥20180228_102025

検査日タグ名 StudyDate

患者IDタグ名 PatientID

データ読み込み 閉じる

検査日 日付の選択 [15] ~ 日付の選択 [15]

検索

内視鏡データ 10 件
病理マッチデータ 5 件
放射線マッチデータ 0 件

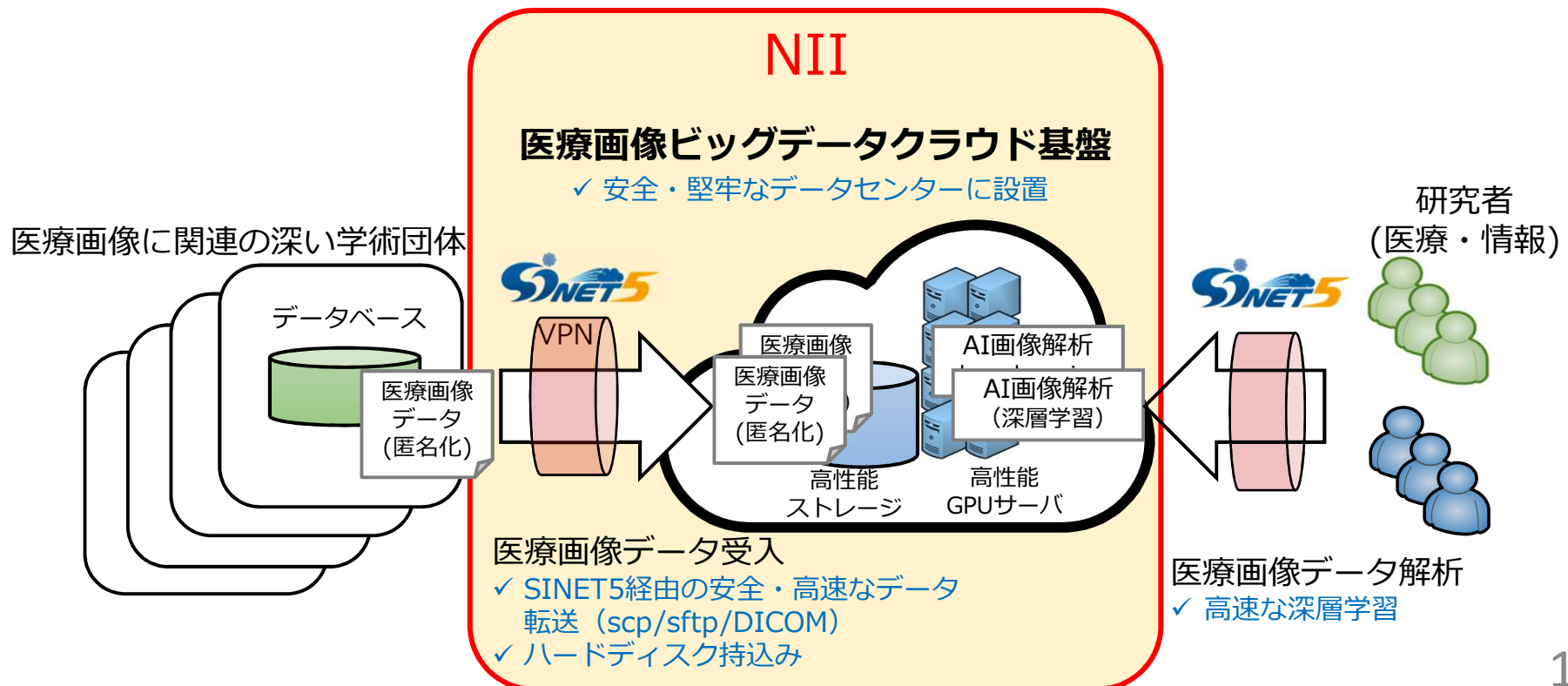
No	検査日	患者ID	病理 マッチ	放射線 マッチ	内視鏡 画像	病理 画像	放射線 画像	ステータス
1	2017/12/01	ID1	●		●	●		
2	2017/12/11	ID11	●		●	●		
3	2017/12/12	ID12			●	●		
4	2017/12/13	ID13			●	●		
5	2017/12/04	ID4	●		●	●		
6	2017/12/05	ID5			●	●		
7	2017/12/06	ID6			●	●		
8	2017/12/07	ID7	●		●	●		
9	2017/12/08	ID8			●	●		
10	2017/12/09	ID9	●		●	●		

平成29年度までの成果 共通プラットフォーム構築

NII

NIIの医療画像データ受入実績（5月現在）

病理	消化器内視鏡	医学放射線
5,000 症例	1,543,246 枚	381,826枚 3万枚／日増加



- ◆ 学会共通課題として、倫理審査承認の遅延が進捗のボトルネックとなっている。
- ◆ AMEDでは学会横断的な取り組みとして、倫理審査有識者を招き、4学会間で以下の内容につきコンセンサスを得た。(H29 第1回調整費)
 - 患者同意取得方法について
 - 医学系指針第5章第12の7 ※¹に該当するとして、**オプトアウトでの収集が可能**である。
 - 倫理審査委員会では**社会的重要性の担保を審査**する。
 - 画像情報の匿名化について
 - 「人を対象とする医学系研究に関する倫理指針ガイダンス」※²に基づき、**適切なマスキングがされていれば匿名化**されている、と解釈できる。
 - これが個人を特定できる情報か否かは、別問題である。
 - あくまで**研究の範疇であれば容認されるという点**に注意が必要である。

※1：スライド22参照

※2：スライド21参照

今後のAI開発ターゲット

◆ 病理学会

下記を用いたがん病変検出のダブルチェック

- 1) 胃生検材料, 2) 大腸生検材料, 3) 婦人科生検材料

◆ 消化器内視鏡学会

1) 胃がんの組織種別も予測できる高精度AI

2) 炎症性腸疾患の病期分類

3) 上部消化管内視鏡の見落とし監視（部位認識AIを応用）

◆ 医学放射線学会

緊急性の高い下記疾患のダブルチェック

- 1) 頭部CT画像を用いた頭蓋内出血, 2) 胸部CTの粗大病変

◆ 眼科学会

眼底写真を用いた下記疾患の診断支援AI

- 1) 緑内障, 2) 糖尿病網膜症, 3) 加齢黄斑変性症

データ共有に係る課題について

今後学会の収集データは、広く公平性を持って提供していくことが望まれる。

しかし、

以下の問題が未解決である。

I. 教師データセット作成に係る障壁

1. メタデータ付与（アノテーション）
2. 匿名化

II. データの第三者提供に係る障壁

1. 個人情報・倫理審査
2. データセキュリティ
3. データ標準化
4. 知的財産権

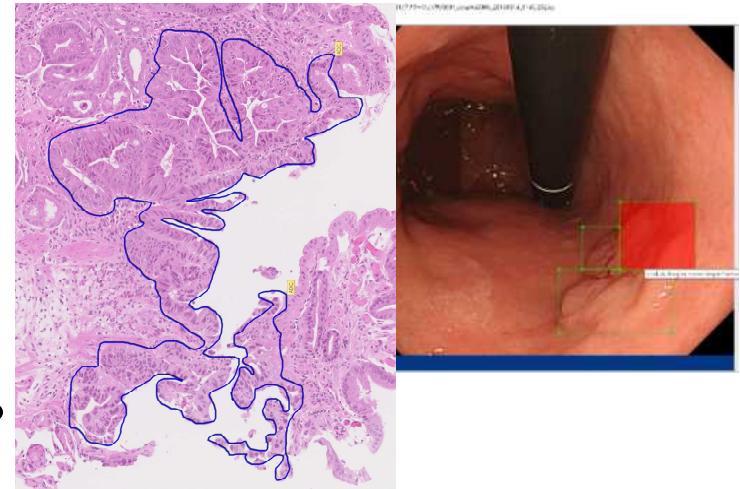
教師データセット作成に係る障壁 メタデータ付与（アノテーション）について

◆ 診断名付与について

- 診断名の正確性の担保はどの領域においても困難である。
- 臨床現場で確定診断名が付与できない領域（消化器内視鏡、放射線、超音波等）がある。

◆ 病変部位マーキングについて

- 臨床現場の通常業務ではない。
（行っているもAI用のアノテーション作業と異なる）
→現場医師に追加の作業負担が生じる。
例）病理画像：専門医の手で1枚数時間
消化器内視鏡画像：熟練者の手で1枚2-3分程度
- 作成するAIの目的によって必要とされる粒度が異なる。



学会ごとにより以下の取り組みが行われている。

- アノテーション補助ツールの開発
- 学会内におけるアノテーションの為のWGの作成
- AIの目的ごとのメタデータの粒度につきNIIと相談

教師データセット作成に係る障壁 匿名化について

「人を対象とする医学系研究に関する倫理指針ガイダンス」によれば、適切なマスキングがされていれば良いと解釈できる。

第5章第12(3)ア(ウ)の「匿名化されているもの」とは、既存資料・情報の中に含まれる記述等のうち、その記述単体で特定の研究対象者を直ちに判別できる記述等を全て削除するような加工がなされているものを指す。ここでいう「記述単体での特定の研究対象者を直ちに判別できる記述等」とは、具体的には例えば、氏名、顔画像、個人識別符号（ゲノムデータ、保険証番号等）が該当する。病名、検査データ等については、その記述等が比較的特異な場合であっても、基本的には「その記述単体で特定の研究者を直ちに判別できる記述等」には該当しないものと取り扱ってよい。

しかし、

- 一部の画像については依然「完全匿名化」の定義が曖昧である。
（再構築可能データと個人の認識が可能なデータとの混乱）
- 完全匿名化により教師データとしての質が劣化する可能性がある。



データの第三者提供に係る障壁 個人情報・倫理審査について

現在は、「インフォームドコンセントの手続き等の簡略化（医学系指針第12の7）に該当するとしてオプトアウトで画像を収集している。

7 インフォームド・コンセントの手続き等の簡略化

(1) 研究者等又は既存試料・情報の提供を行う者は、次に掲げる要件の全てに該当する研究を実施しようとする場合には、研究機関の長の許可を受けた研究計画書に定めるところにより、1及び2の規定※による手続きの一部を簡略化することができる。

- ① 研究の実施に侵襲（軽微な侵襲を除く。）を伴わないこと。
- ② 1及び2の規定※による手続きを簡略化することが、研究対象者の不利益とならないこと。
- ③ 1及び2の規定※による手続きを簡略化しなければ、研究の実施が困難であり、又は研究の価値を著しく損ねること。

④ **社会的に重要性が高い研究と認められるものであること。**

※ 1及び2の規定：同章「インフォームド・コンセントを受ける手続き等」・「研究計画書の変更について」を指す。

しかし、

- 医療情報部、倫理審査委員会等の理解を得られないことがある。
 - 本当に実施困難か、社会的に重要性が高いか
 - 解釈が困難な中での拡大解釈
- 企業等へのデータ提供を想定した場合、現行方式では成り立たない

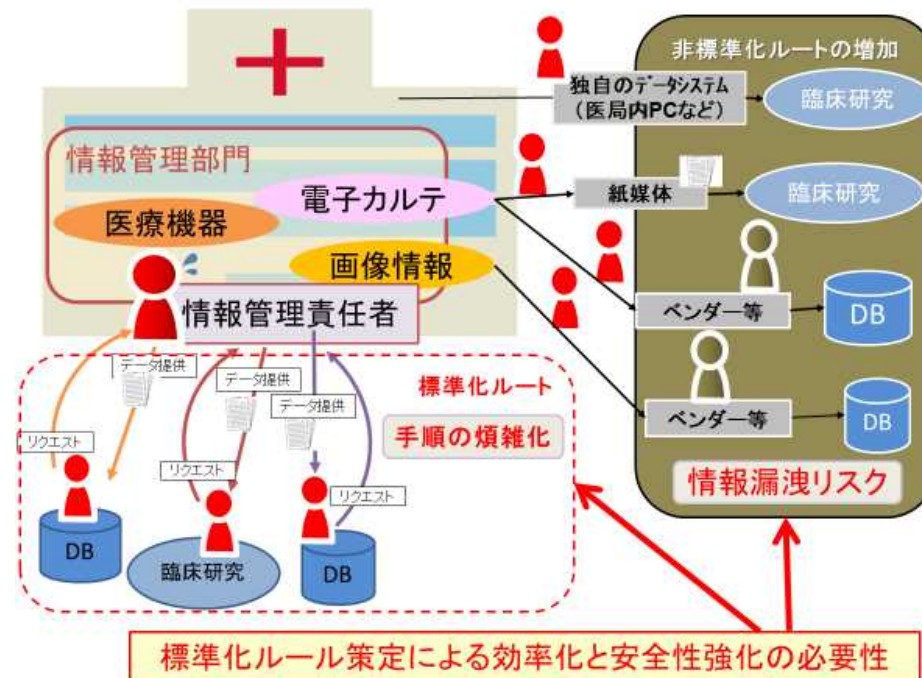
データの第三者提供に係る障壁 データセキュリティについて

倫理審査とは別に、情報管理者による情報漏えいの監督も必要である。

しかし、

- 情報管理者の負担増
- ガイドライン等の不在

により、説明・同意の取得が困難な場合がある。



AMEDでは本年度より、医療情報学会主導の研究開発を委託

データの第三者提供に係る障壁 データの標準化について

データ共有のためには、少なくとも以下においてベンダーニュートラルな標準化が必要である。

- データフォーマット
- 規格
- 格納形式
- HIS※からの抽出方法
- 付帯情報（項目・用語・書式）
- 読影レポートのフォーマットや用語

しかし、

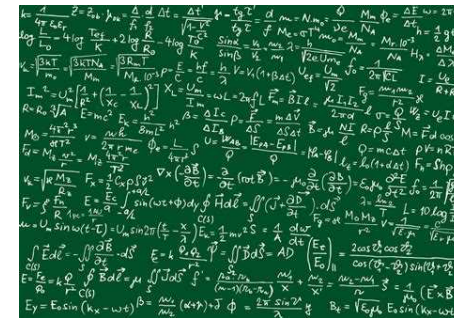
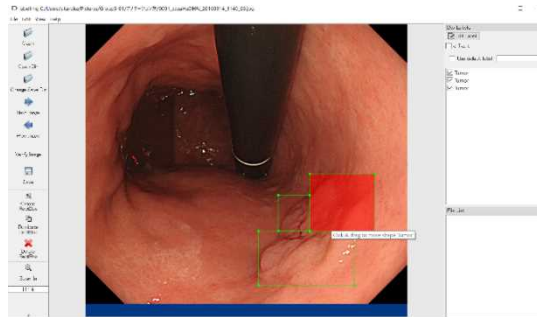
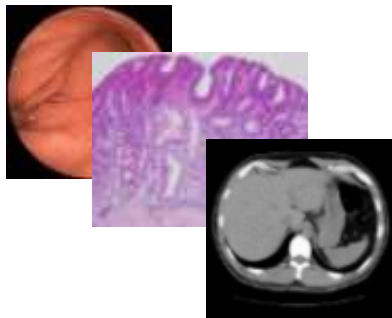
- 全てを統一するのは困難である。
（MRI画像のテスラ数、超音波画像の周波数など）
- 学会内である程度企画を統一しても、現場全体の統一は困難である。

※HIS : Hospital Information System

データの第三者提供に係る障壁 知的財産権について

広範かつ公正な共有が求められるデータには以下がある。

- 学会等の所有する画像データ
- 学会等の作成したアノテーション済み教師データセット
- 学会等が開発するアルゴリズム



しかし、

- データセットが無償提供されることにより、アノテーターのモチベーションが下がる可能性がある。
- 収集画像データは学会の所有物であるが、データセットの所有権については不明瞭である。
- ある研究者が類似したアルゴリズムの特許権を取ってしまうことで、研究に支障が生じる可能性がある。

課題解決に有用となり得る指針等（案）

◆倫理審査に係る課題について

- 医学研究・診療分野での医療情報の利活用の形態は多様化している。
- 医療情報を利活用した研究促進・社会還元を図るには、法的に容認される範囲、研究形態等の更なる明確化が望まれている。

◆標準化に係る課題について

- DICOM規格と同様に、画像データ全体に共通した収集・格納方法などの規格が望まれる。

◆第三者提供に係る課題について

- 同意取得や画像匿名化につき、法律上許容できる具体例の明示が望まれる。
- 医療等AIのデータの所有や知財の在り方（アノテーション済み教師データの所有権、AIの特許や知的財産権、知的財産の管理方法、研究段階での留意事項など）についての指針があることが望ましい。

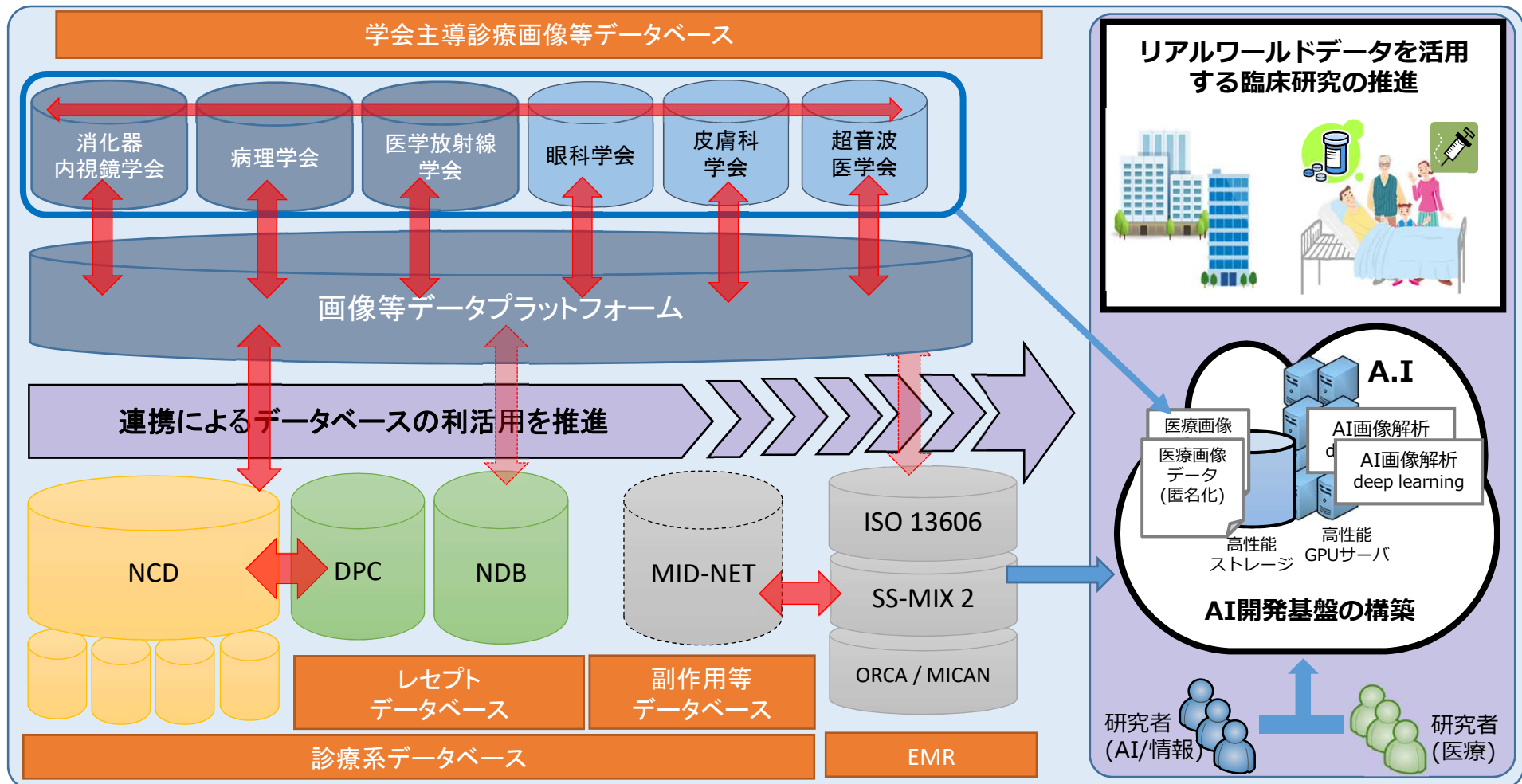
◆公益性に係る課題について

- 優先的な開発を推奨するAI開発分野につき、国の指針を策定することが望ましい。

AI解析等も見据えた医療ビッグデータプラットフォームの構築

参考資料

- 医療分野研究開発推進計画（平成29年2月健康・医療戦略本部決定）では、「**医療・介護等のデジタルデータの利活用基盤の構築を進めるとともに、革新的な人工知能の基盤技術を構築し、収集されたビッグデータを基に人工知能技術を活用することで、診療支援や新たな医薬品・医療技術の創出に資する研究開発を進める。**」とされている。
- また、「保健医療2035」では人々が世界最高水準の健康、医療を享受でき、安心、満足、納得を得ることができる持続可能な保険医療システムを構築し、我が国及び世界の繁栄に貢献することを目標とされており、保健医療分野におけるICT活用推進懇談会では「2020年までに産官学の多様なニーズに応じて、保険医療データを目的別に収集・囲う（匿名化等）提供できる「データ利活用プラットフォーム」（仮称）を整備」することとされている。さらに、厚生労働省において開催されているAI活用推進懇談会においては、**画像データベースを早急に構築**すること等の必要性について指摘がある。このため、2017年度中に画像DBの構築・拡大を行い同時に各種連携の課題を抽出するとともに世界的なAI開発競争の状況も踏まえると、AI活用の検討を加速する必要がある、2018年度には各学会主導画像DBの完成度を高め、DB間連携及びAI活用研究の実装に繋げていく必要がある。



診療画像等データベースプラットフォームの構築(現状と今後) - ICT技術や人工知能(AI)等による利活用を見据えたインフラの整備 -

