

学位論文審査の結果の要旨

審査区分 課・論	第688号	氏名	中 沼 寛 明
審査委員会委員	主査氏名	下 村 剛	印
	副査氏名	藤 木 穂	藤木
	副査氏名	安 藤 忠 助	安藤
論文題目			
<p>An intraoperative artificial intelligence system identifying anatomical landmarks for laparoscopic cholecystectomy: a prospective clinical feasibility trial (J-SUMMIT-C-01) (人工知能による腹腔鏡下胆嚢摘出術中の解剖学的ランドマーク教示システムの開発 前向き臨床性能試験; J-SUMMIT-C-01)</p>			
論文掲載雑誌名			
Surgical Endoscopy			
論文要旨			
<p>本研究の目的は、術中リアルタイムランドマーク教示 AI システムの臨床現場における性能試験を行い、AI システムの解剖学的ランドマークの教示性能の客観的な評価方法を確立することである。Smart Endoscopic Surgery (SES)は、YOLOv3 を用いて構築し、76 例の腹腔鏡下胆嚢摘出術 (LC) ビデオから抽出した胆嚢頸部 Calot 三角領域の 2339 の内視鏡画像上の解剖学的構造物に対し、熟練外科医の見立てをそれぞれ、総胆管 (LM-CBD)、胆嚢管 (LM-CD)、肝 S4 の下縁 (LM-S4)、ルビエレ溝 (LM-RS) としてタグ付けし、教師データとした。2019 年 9 月から 2020 年 3 月に大分大学医学部附属病院で施行した LC10 症例において単施設前向き臨床性能試験を行い、主評価項目である LM-CBD、副評価項目である LM-CD、LM-S4、LM-RS それぞれのランドマーク教示精度の評価を行った。AI 教示画像の正確性についてはループリックに基づく 5 段階評価を行った。次に外部評価委員 3 名が、臨床性能試験時の内視鏡画像に対してアノテーションを行い、これらを AI の教示画像と比較し、DICE 係数にて AI 教示の定量的評価を行った。</p> <p>外部評価委員による AI 教示画像の正確性に関する 5 段階評価は、平均値で、LM-CBD、LM-CD、LM-S4、LM-RS の順に、4.2 ± 0.8、3.6 ± 0.8、4.9 ± 0.2、4.1 ± 1.0 であった。外部評価委員同士の DICE 係数の平均値は、LM-CBD、LM-CD、LM-S4、LM-RS の順に 0.72、0.49、0.46、0.66 であり、AI が教示した解剖学的ランドマークのアノテーション領域と、各外部評価委員が示したアノテーション領域との DICE 係数に有意な差は認められなかった。主評価項目である CBD の AI 教示とエキスパートとの間の DICE 係数は、エキスパート間の DICE 係数の平均値とほぼ同等であった。また、LM-CBD の AI 教示とエキスパートとの間の DICE 係数は、エキスパート間の DICE 係数とほぼ同等であった。DICE 係数は、感度と特異度の調和平均であり、AI の評価尺度として一般的に用いられる手法の一つである。今回の実機検証において、外部評価委員の 5 段階評価が高評価であった症例において、CBD の DICE 係数が 0.72 と高得点であったことから、DICE 係数が AI によるランドマーク教示システムを評価する一つの指標となり得ると考えられた。</p> <p>本研究は、術中利用を前提とする AI システムの臨床性能試験の実施と外部評価者による AI システムの評価を試みた世界初の報告であり、ループリックを用いた 5 段階評価と DICE 係数を組み合わせた評価は、AI システムの臨床的な評価法として有用であることを示した。このため、審査員の合議により本論文は学位論文に値するものと判定した。</p>			

最終試験
の結果の要旨
~~学力の確認~~

審査区分 ① 課・論	第688号	氏名	中 沼 寛 明
審 査 委 員 会 委 員	主査氏名	下 村 剛	①
	副査氏名	藤 木 稔	②
	副査氏名	安 藤 忠 助	③
<p>学位申請者は本論文の公開発表を行い、各審査委員から研究の目的、方法、結果、考察について以下の質問を受けた。</p> <ol style="list-style-type: none"> 1. この研究の教師データの数は妥当ですか。 2. DICEを使った理由と他の評価方法について述べよ。 3. AIに教え込む画像の準備は手間と負担の大きいのが、負担を減らす改善策はについて述べよ。 4. 本研究の成果物はいかなる外科医を対象と想定するか述べよ。 5. AIは外科医の人為的過誤を凌駕するか考察せよ。 6. 強い炎症がある場合の対処法について述べよ。 7. 術前の画像データを利用する方法について述べよ。 8. 今後どのような疾患や臓器に応用可能と考えているか述べよ。 9. 稟事承認を見据えた戦略を概説せよ。 10. 暗黙知には触覚も重要と思われるが、AIに活用する方法はあるか。 <p style="margin-top: 20px;">これらの質疑に対して、申請者は概ね適切に回答した。よって審査委員の合議の結果、申請者は学位取得有資格者と認定した。</p>			

(注) 不要の文字は2本線で抹消すること。

学 位 論 文 要 旨

氏名 中沼 寛明

論 文 題 目

An intraoperative artificial intelligence system identifying anatomical landmarks for laparoscopic cholecystectomy: a prospective clinical feasibility trial (J-SUMMIT-C-01)

(人工知能による腹腔鏡下胆嚢摘出術中の解剖学的ランドマーク教示システムの開発
前向き臨床性能試験; J-SUMMIT-C-01)

要 旨

〔 緒言 〕 深層学習 (Deep learning) による画像認識技術の向上により、人工知能 (AI) を用いた医療機器開発の試みが、外科手術の領域でも始まっている。我々は、熟練外科医が術中の動的環境下における判断基準として用いる解剖学的ランドマークを AI が教示する医療システム (Smart Endoscopic Surgery, 以下 SES) を構想し、その実現に取り組んできた。しかし、AI による手術支援システムを臨床現場で検証した報告はなく、評価方法も確立していない。本研究の目的は、我々が開発した術中リアルタイムランドマーク教示 AI システムの臨床現場における性能試験を行い、AI システムの解剖学的ランドマークの教示性能を客観的に評価し、評価方法を確立することである。

〔 研究対象及び方法 〕 SES は、深層学習に基づく対象検出アルゴリズムである YOLOv3 を用いて構築した。76 例の腹腔鏡下胆嚢摘出術 (LC) ビデオから抽出した胆嚢頸部 Calot 三角領域の 2339 の内視鏡画像上の解剖学的構造物に対し、熟練外科医の見立てをそれぞれ、総胆管 (LM-CBD)、胆嚢管 (LM-CD)、肝 S4 の下縁 (LM-S4)、ルビエレ溝 (LM-RS) としてタグ付けし、教師データとした。

2019年9月から2020年3月に大分大学医学部附属病院 IRB 承認下に、当科で施行した LC10 症例において単施設前向き臨床性能試験を行った。臨床性能試験実施後に外部評価委員会を設置し、主評価項目である LM-CBD、副評価項目である LM-CD、LM-S4、LM-RS それぞれのランドマーク教示精度の評価を行った。まず、AI 教示画像の正確性について、ルーブリックに基づく 5 段階評価を行った。次に外部評価委員 3 名が、臨床性能試験時の内視鏡画像に対してアノテーションを行い、これらを AI の教示画像と比較し、DICE 係数を算出し、AI 教示の定量的評価を行った。

〔結果〕 3 人の外部評価委員による AI 教示画像の正確性に関する 5 段階評価は、10 症例の平均値で、LM-CBD、LM-CD、LM-S4、LM-RS の順に、 4.2 ± 0.8 、 3.6 ± 0.8 、 4.9 ± 0.2 、 4.1 ± 1.0 であった。外部評価委員同士の DICE 係数の平均値は、LM-CBD、LM-CD、LM-S4、LM-RS の順に 0.72、0.49、0.46、0.66 であり、AI が教示した解剖学的ランドマークのアノテーション領域と、各外部評価委員が示したアノテーション領域との DICE 係数に有意な差は認められなかった。主評価項目である CBD の AI 教示とエキスパートとの間の DICE 係数は、エキスパート間の DICE 係数の平均値とほぼ同等であった。

〔考察〕 AI による解剖学的ランドマークの術中教示システムを構築し、10 症例の LC において臨床の現場での実機検証試験を行った。LM-CBD の AI 教示とエキスパートとの間の DICE 係数は、エキスパート間の DICE 係数とほぼ同等であった。DICE 係数は、感度と特異度の調和平均であり、AI の評価尺度として一般的に用いられる手法の一つである。今回の実機検証において、外部評価委員の 5 段階評価が高評価であった症例において、CBD の DICE 係数が 0.72 と高得点であったことから、DICE 係数が AI によるランドマーク教示システムを評価する一つの指標となり得ると考えられた。

〔結語〕 今回我々は、開発した SES の臨床現場における性能試験を行い、解剖学的ランドマークの教示精度を評価した。術中利用を前提とする AI システムの臨床性能試験の実施と、外部評価者による AI システムの評価に関する試みは世界初であり、ルーブリックを用いた 5 段階評価と DICE 係数を組み合わせた評価は、AI システムの臨床的な評価法として有用であることが示唆された。今後、AI システムの実用化に向けて、手術シーン認識技術や炎症症例などの高難度教師データを AI に組み込み、更なるランドマーク教示精度を向上させ、社会実装を実現したい。