

内科におけるコンピュータ利用診断支援システムによる診断改善

栗山 裕^a, 曾田 祐民^a, 矢野 愛華^a, 安田 英己^b,
石井 修^c, 斉尾 武郎^{d,e}, 鳥越 恵治郎^{c*}, 上田 剛士^f,
志水 太郎^g, 徳田 安春^h

^a岡山大学医学部医学科, ^b医療法人安田内科医院, ^c医療法人恵真会 鳥越医院, ^dSMBC日興証券 健康管理室,
^eフジ虎ノ門整形外科病院 内科・精神科, ^f洛和会丸太町病院 救急総合診療科, ^g獨協医科大学 総合診療医学,
^h臨床研修病院群プロジェクト 群島沖繩臨床研修センター

Better diagnostic performance using computer-assisted diagnostic support systems in internal medicine

Yutaka Kuriyama^a, Yumi Sota^a, Aika Yano^a, Hideki Yasuda^b,
Osamu Ishii^c, Takeo Saio^{d,e}, Keijiro Torigoe^{c*}, Takeshi Ueda^f,
Tarou Shimizu^g, Yasuharu Tokuda^h

^aOkayama University Medical School, Okayama 700-8558, Japan, ^bYasuda Clinic of Internal Medicine, Okayama 700-0862, Japan,
^cTorigoe-iin, Okayama 715-0025, Japan, ^dHealth Management Office, SMBC Nikko Securities Inc., Tokyo 100-8325, Japan,
^eDepartment of Internal Medicine and Psychiatry, Fuji Toranomon Orthopedic Hospital, Shizuoka 412-0045, Japan,
^fEmergency and General Internal Medicine, Rakuwakai Marutamachi Hospital, Kyoto 604-8401, Japan,
^gDepartment of Diagnostic and Generalist Medicine, Dokkyo Medical University, Tochigi 321-0293, Japan,
^hMuribushi Okinawa for Teaching Hospitals, Okinawa 901-2132, Japan

The recent application of artificial intelligence (AI) to clinical medicine has confirmed the usefulness of AI for diagnostic imaging, histopathological examinations, and dermatologic screening. Clinical decision support systems are another promising area to which AI could contribute toward better clinical decisions. We have developed computer-assisted diagnostic support systems to reduce human diagnostic errors such as delayed diagnoses, misdiagnoses, and overdiagnoses. Our three Diagnosis Reminder (DR) systems include two AI systems that use machine learning in their diagnosis algorithms. Here, we compared the diagnostic accuracy of a DR-supported group with that of an unassisted physicians group, using three difficult patient cases provided by experts in general medicine.

Our analyses revealed that the three AI diagnostic systems could not provide accurate differential diagnoses up to top 10 in all three patient cases because of incomplete data inputs for machine learning. However, the first DR system, which was developed by an experienced diagnostician over the last 35 years, showed very useful performance in reducing human diagnostic errors when it was used by an expert physician. The use of AI diagnostic systems by knowledgeable physicians will lead to better diagnostic performance. We also discuss the current scenario, future challenges, and prospects for AI diagnostic systems herein.

キーワード：AI 診断システム (AI diagnostic systems), 診断思い出し (diagnosis reminder), 機械学習 (machine learning), 診断エラー (human diagnostic errors)

緒 言

近年コンピュータにおいてCPU・GPU・メモリなどのハードの性能向上に加え、Google社のTensorflowなどの機械学習用ソフトウェアが公開され、誰でも機械学習が可能になったことにより第三次AI(artificial intelligence; 人工知能)ブームが到来している。

ちなみに第一次AIブームとは1950年代後半から1960年代で、多くの研究者が推論と探索によって問題を解く研究を行った。ただ推論と探索はルールとゴールが厳密に定義された枠組のなかでしか利用できなく、またコンピュータの性能も低かったため現実世界では役に立たなかった。第二次AIブームは1980年代で、コンピュータに専門家が持つ知識を学習させて問題を解決させるエキスパートシステムと呼ばれるシステムが開発された。しかし医師などの専門家がもつ知識を記述し、コンピュータに教える作業は非常に煩雑であり、またコンピュータの能力自体もまだ十分でなかったため、実用に耐えられるAIシステムの開発意

平成30年9月7日受理

*〒715-0025 岡山県井原市笹賀町2-21-3
電話：086-663-1656 FAX：086-663-1658
E-mail：torigoe@cc.okayama-u.ac.jp

欲は急速にしばんでいった。

そして2000年代に入って、機械学習という方法論を用いて、コンピュータに多量のデータを与え、コンピュータ自身がデータから学習するという試みが発展した。そしてコンピュータの性能向上、インターネットの普及、クラウドによる膨大なデータの蓄積により、実用的なシステムの開発が可能になった。これが現在の第三次 AI ブームの始まりである。

2018年の現在、医療界において特に放射線・皮膚・病理学においては、それぞれ画像診断、皮疹診断、病理診断の AI が開発されている。皮膚科領域では専門家と AI に画像診断をさせたところ、結果にほぼ差がない¹⁾。AI のほうが多くの皮膚科医より勝っている²⁾といった報告もなされている。

筆者らは内科診断において臨床症状・検査所見などに対して、可能性の高い疾患をヒト（医師）に提示することを目的とした AI を利用することで、医師の診断エラーを減らすことができるのではないかと考えている。鳥越とプログラマの石井は AI 診断システムとして diagnosis reminder（以下、T-DR）を開発してきた³⁻⁶⁾。最近では石井、栗山により T-DR をさらに発展させるために、アルゴリズムにニューラルネットワークを採用した機械学習を用いて、疾患を学習させた I-DR（Ishii-DR）・K-DR（Kuriyama-DR）が開発されている。これら現状の AI 診断システムを確定診断に苦慮した症例に対して適用し、その結果から内科診断における AI 活用の有用性・問題点・今後の可能性について述べる。

方 法

2018年3月31日に岡山県井原市において「総合診断 ヒト・マシン対決」と銘打って総合診断研究会が開催された。そこでは、現在の AI 診断システムがヒトの臨床推論に対してどこまで迫れるのか、どのような問題点があるかが検討された。

検討症例は共著者であり総合診療専門医の上田・志水・徳田が、それぞれが経験した、最終診断に対して典型的な症状を示しておらず確定診断に苦慮した症例1例ずつの3症例を提示した。聴衆として学生、初期・後期臨床研修医、卒後10年前後の総合臨床医、卒後15年以上の内科医約60名が参加し症例検討を行った。症例の要約を表1（出題：上田）・表2（出題：志水）・表3（出題：徳田）に示す。

同時に3症例（症例1：イソニアジド中毒、症例2：気管支カルチノイド腫瘍、症例3：甲状腺クリーゼ）に対して3種類の AI 診断システム（T-DR・I-DR・K-DR）を適用した。AI 診断システムに入力するデータ（陽性症状・陽性所見）は、鳥越が各症例の現病歴、検査値をもとに設定した。

各 AI 診断システムの特徴は下記のとおりである。

T-DR は鳥越が約35年かけて約3,000の疾患とそれぞれの自他覚症状・検査所見を集め、それを疾患の基本的な固有データ群として予め登録したツールである。疾患検索アルゴリズムに機械学習は使われておらず、鳥越自身がシステム内部の係数等を調整することで作成したシステムである^{3,5)}。I-DR は T-DR の出力結果を入力データに対する結果とし

表1 症例1サマリ

症例1：80歳女性 主訴 興奮 現病歴：1年前に関節リウマチを発症しメトトレキサートを開始した。4か月前に器質性肺炎をきたしステロイドパルスを行い、人工呼吸器管理・気管切開となった。その後、呼吸器離脱しプレドニゾロンは45mg/日より漸減していた。3週間ほど前より興奮状態となった。ステロイド精神病を疑いプレドニゾロン17.5mg/日から10mg/日に減量しクエチアピンを開始したが精神状態は増悪し、ほとんど睡眠をとることもなく四六時中興奮している。発熱、頭痛、嘔気、幻覚はない。 既往歴：関節リウマチ、リウマチ関連間質性肺炎（器質性肺炎）、潜在性結核感染症（リウマチ関連間質性肺炎と同時に治療開始）、冠縮性狭心症。 内服薬：プレドニゾロン 10mg、アスピリン 100mg、ST 合剤（スルファメトキサゾール 200mg、トリメトプリム 40mg）、アルファカルシドール 0.25 μ g、ジルチアゼム 100mg、クエチアピン 100mg、イソニアジド 200mg、ピリドキシン塩酸塩 30mg、リセドロン酸ナトリウム 75mg（月1回）。 身体所見：体温 37.2 $^{\circ}$ C、血圧 189/118mmHg、脈拍 115回/分・整、呼吸数 16回/分、SpO ₂ 96%（室内気）、GCS：E4V4M5、一般身体所見に特記すべき所見なし。神経脱落所見はない。高次脳機能は興奮のため評価困難だが失語や失行は認めない。 血液検査：白血球数 13,800/ μ L、CRP 4.63mg/dL 以外に特記すべき所見なし。甲状腺機能正常。抗核抗体・甲状腺自己抗体・STS・TPHA・HIV 抗体・クリプトコッカス抗原は陰性。 胸部 CT：3か月前と比較して右中葉、右肺底部の浸潤影は増強、左肺底部の浸潤影は消退傾向。 頭部 MRI：特記すべき所見なし。 髄液：特記すべき所見なし。 最終診断：イソニアジド中毒

て利用しながら、T-DRの基本的な固有データ群を機械学習させたシステムである。K-DRは鳥越が収集した症例を参考に危険因子・疾患分類などを固有データ群として加え

てやや詳細に収集した疾患データ群を機械学習させたシステムである。

表2 症例2サマリ

<p>症例2：58歳女性 主訴 8か月続く慢性の咳 現病歴：特に既往のない58歳女性。約8か月前から時々痰が絡む空咳が出るようになった。ここまで長いのは初めてとのことだった。咳は朝夜の変化・体温・気温・季節・運動の有無・移動する場所・立位臥位の変化には関係がない。また、口の苦い感じ・胸やけ・食事中のつかえ感・げっぷ、血痰、食欲低下、体重減少はない。1年前に大道路に面したマンションの3階に引越したが、排気ガスは気にならない。半年前のある日から数日で痰が増えてきて息苦しさ・熱っぽさを感じ内科を受診したところ胸部レントゲンで肺炎と言われ、抗菌薬を1週間内服して熱と咳は2週で治まった。しかし数日後から以前と同じ咳が出現し再び熱が出た。再度同じ内科を受診し肺炎の再燃と言われ、同じ抗菌薬を1週間内服して解熱し痰も減ってきた。しかし咳だけがぶり返し持続した。2か月後の経過観察でレントゲンでは肺炎は軽快したと言われた。採血も問題なかったが、気管支喘息があるかもしれないと言われ、高容量ステロイドと長時間作用型β刺激薬の吸入を処方され1か月ほど様子を見た。抗ヒスタミン剤とPPIも追加したが2週間でも効果がなかった。咳自体は増加傾向ということで近医から紹介受診となった。 既往歴・内服歴：上記。アレルギー：花粉症、春に涙・鼻水・咳が出る。家族歴：なし。 生活歴：62歳の夫とふたり暮らし。大卒後30歳の時まで印刷会社勤務で、その後結婚を契機に退職し専業主婦となった。35歳時に夫のグラフィックデザインの仕事を手伝いつつ自宅兼事務所で仕事を始めた。同年妊娠、出産し男児をもうける。息子は出産、成育歴共に問題ない。生活環境は1年前、結婚以後長らく住んでいた実家を離れ、マンションに移った。マンションは築数年のコンクリートの2LDK。卸売り店でパートの店員を始めた。仕事は週4で9時～15時で軽作業のみ、症状は仕事のオンオフで変わらない。夫による受動喫煙はあるが、10年前から夫はベランダで吸うようにしている。飲酒は機会飲酒程。お酒を飲むと咳が強くなる気がして最近では飲まなくなった。 身体視察：やせ型、呼吸促拍はない。話しているとき咳き込みは強くない。血圧120/78mmHg・脈拍80/分(整)・呼吸数15回/分・SpO2 98%(室内気)・体温36.5度。眼底は正常所見、耳鏡は正常。鼻粘膜は正常で、咽頭後壁は後鼻漏の所見もなし。口腔内潰瘍なし。頸静脈怒張なし。甲状腺腫大なし。頸部・腋窩リンパ節腫大なし。心音：S1, S2は正常で過剰心音なし。心雑音、心膜摩擦音なし。呼吸音：聴診上右中葉領域にかすかな wheeze があり、強制呼気でやや増強、臥位で変わらず、咳ばらいで減弱する。四肢：下腿浮腫なし・脈拍触知良好・ばち指なし・チアノーゼなし・爪周囲炎無し。関節炎の所見なし。皮膚：皮疹なし・白斑なし。頭髮も異常なし。 胸部レントゲン：聴診での該当部位に明らかな腫瘍性病変はなし。縦隔・心陰影の拡大無し。胸水なし。肺野に空洞無し。浸潤影なし。造影 HRCT：右中葉枝からB4で石灰化を伴う軽度の造影効果を有する腫瘍と近傍の気管支壁の肥厚、気管支拡張像を認めた。明らかな腫瘍の壁外進展の所見はなかった。 気管支内視鏡：気道壁に沿った無茎性ポリープ様隆起病変あり、生検結果はカルチノイド腫瘍と判明した。尿中5-HIAAは正常範囲。転移検索は陰性で、呼吸器外科で核出術を行った。 最終診断：気管支カルチノイド腫瘍(Stage I)</p>
--

表3 症例3サマリ

<p>症例3：47歳女性，主婦 主訴 4日前からの心窩部痛 現病歴：4日前から徐々に発症する心窩部の痛みで歩行来院。痛みの程度は10分の10で、鈍い痛み。当初は間欠的な痛みであったが前日からは持続痛となった。食事摂取は3分の1まで低下していた。同時期に泥状便1日5回あり。血便や不正性器出血は無し。昨日より38度程度の発熱あり。渡航歴、動物の生肉摂取、月経異常、飲酒、喫煙、などなし。普段の内服薬なし健康診断は最近数年受けていない。 身体所見：全身外観は病的。血圧 95/52mmHg, 心拍数毎分 160 整, 呼吸数毎分 19, 体温摂氏 38.0度。頭頸部異常なし。甲状腺腫大なし。心音は整で心雑音なし。呼吸音は清で副雑音なし。腹部は平坦で軟らかいが、腹部全体に軽度の圧痛あり、最強点は心窩部。皮疹なし。末梢冷感なし。神経学的に異常所見なし。 検査結果：血算・血液生化学：白血球数 15,900/μL(白血球分画では好中球増多あり), ヘモグロビン 13.0 g/dL, 血小板 30万/μL, BUN 40mg/dL, Cr 3.1mg/dL, Na 145mEq/L, K 4.1mEq/L, Cl 100mEq/L, Ca 12.5mg/dL(補正Ca 12.6mg/dL), Mg・P 正常範囲内, Alb 3.9g/dL, AST・ALT・ALP・γGTP・CPK・LDH 正常範囲内, CRP 0.5mg/dL, トロポニンT 上昇なし, リパーゼ 上昇なし, Dダイマー 上昇なし。 検尿：異常なし。 腹部超音波：明らかな異常所見を認めず。 心エコー：構造的に明らかな異常所見を認めず。 最終診断：甲状腺クリーゼ</p>
--

結 果

1. AI 診断

各症例の AI 診断システムへの入力データは表 4 の通りである。

次に表 5 に各症例の AI 診断システムの出力結果の上位 5 疾患を示す。また各システムが正解を何位に挙げたかを表 6 に示す。T-DR では絞込み検索が可能^{4,6)}で、結果を参考として記載した。絞込み検索の詳細は参考文献などに委ねるが、ここでは各症例において、それぞれの入力データ

の全てをもって絞り込んだ。

全ての症例で AI 診断システムは正解となる疾患をトップ10位以内の上位に提示できていない。

2. ヒトによる症例検討の結果

症例検討におけるヒトの診断結果を、その診断を挙げた医師が多い順に述べる。症例 1 では多剤投与されていたために何らかの薬剤が原因ではないかと考える医師が最も多く、数人がイソニアジド中毒を疑った。症例 2 では非結核性抗酸菌症、肺炎、結核の順に疑われた。気管支カルチノイド腫瘍を疑う意見は出なかった。症例 3 は急性心筋梗

表 4 各症例の入力データ

症例	それぞれの DR に入力した症状・検査所見・危険因子など
1. イソニアジド中毒	興奮, 不眠症, 食欲不振, 発熱, 脱水症, ステロイド使用中, 薬物副作用 (疑), 頻脈, 高血圧, LDH 増加, CRP 陽性, 白血球増多, 胸部異常陰影
2. 気管支カルチノイド腫瘍	咳, 痰・喀痰増加, 呼吸困難, 喘鳴, ほてり, 胸部異常陰影, 自己免疫疾患 (治療中), 腫瘍
3. 甲状腺クリーゼ	腹痛, 腹部圧痛, 嘔気, 嘔吐, 食欲不振, 下痢, 発熱, 頻脈, 血圧低下, 多汗, 頻呼吸, 心電図異常, 白血球増多, 尿素窒素高値, 血清クレアチニン増加, 高カルシウム血症, 低アルブミン血症

表 5 各症例の AI 診断結果

症例 1 : イソニアジド中毒

System	Most likely	2nd	3rd	4th	5th	
T-DR	一次検索	サリチル酸中毒症	急性呼吸促進症候群	敗血症／菌血症	急性肺炎・胸膜炎	(注射) 薬物副作用
	絞り込み	敗血症／菌血症	急性肺炎・胸膜炎	(注射) 薬物副作用	糖尿病性昏睡	クリプトコッカス症
I-DR	サリチル酸中毒症	急性呼吸促進症候群	急性肺炎・胸膜炎	急性心外膜炎	結節性多発動脈炎	
K-DR	急性肺炎・胸膜炎	EB ウイルス感染症	非結核性抗酸菌症	急性肺梗塞症	敗血症 / 菌血症	

症例 2 : 気管支カルチノイド腫瘍

System	Most likely	2nd	3rd	4th	5th	
T-DR	一次検索	気管内腫瘍	特発性間質性肺炎	薬剤誘起性肺臓炎・誤嚥性肺臓炎	アスペルギルス症	アレルギー性気管支肺真菌症
	絞り込み	ノカルジア症	好酸球性多発血管炎性肉芽腫症 (EGPA)	縦隔腫瘍	肺癌または (主) 気管支癌	上大静脈症候群
I-DR	気管内腫瘍	薬剤誘起性肺臓炎・誤嚥性肺臓炎	特発性間質性肺炎	肺癌または (主) 気管支癌	好酸球性多発血管炎性肉芽腫症	
K-DR	非結核性抗酸菌症	原因不明の器質化肺炎	気管支喘息・アスピリン喘息	上大静脈症候群	クリプトコッカス症	

症例 3 : 甲状腺クリーゼ

System	Most likely	2nd	3rd	4th	5th	
T-DR	一次検索	十二指腸または胃穿孔	内臓破裂の可能性	非閉塞性腸間膜梗塞症	多臓器不全症候群	結核性腹膜炎
	絞り込み	内臓破裂の可能性	結核性腹膜炎	心筋炎	甲状腺クリーゼ	アジソン病
I-DR	十二指腸または胃穿孔	内臓破裂の可能性	多臓器不全症候群	非閉塞性腸間膜梗塞症	腸間膜動脈血栓塞栓症	
K-DR	毒素性ショック症候群	急性腸間膜リンパ節炎	急性虫垂 (憩室) 炎	急性心筋梗塞症	慢性腸管破裂による膿瘍	

表6 各症例において正解をあげた順位

System	症例1※	症例2	症例3
T-DR	5	69	29
I-DR	9	68	28
K-DR	138	50	96

※症例1についてはイソニアジド中毒を学習させていないため最終診断を提示できていない。ここに示した数字は「(注射)薬物副作用」の順位

塞，急性心筋炎，甲状腺機能亢進症が疑われた。

考 察

本研究では筆者らが開発した3つのAI診断システムを総合診療の専門家が過去に経験した3つの診断に苦慮した症例に適用し，その結果を検討した。

結果としては，どの症例においてもAI診断システムは最終診断を上位で提示することはできなかった。その原因としては症例1では基礎疾患としてイソニアジド中毒を学習させていなかったことにある。症例2，3についてはそれぞれについて学習させたデータ数が十分でなかったと思われる。さらに表6からはK-DRの能力不足が明らかである。これはK-DRでマシン・ラーニングを始めて約1年経過した時期の試練であり，ラーニング・データの不足を示唆するものと考えられる。ちなみにI-DRはマシン・ラーニングを始めて約3年経過している。

しかし症例1については広く薬物副作用として捉えるとT-DR，I-DRでは10位以内に挙げられており，症例2についても広く気管内腫瘍として捉えるとT-DR，I-DRでは1位で挙げられている。症例3でもこの両者は30位以内に正診を挙げており，今回のような非特異的な症状を示す症例に対して筆者らが開発したAI診断システムは正解までは出せなくても正診に近い疾患名を提示することができると考えている。

次に医師と筆者らが開発したAIの臨床推論の方法について考える。医師の用いる診断論理として，古典的にはEdmund A. Murphyの示した4つの方法（① Gestalt（ゲシュタルト的接近法（パターン認識）），② exhaustive enquiry（総当たり接近法），③ multiple branching process（多分枝選択的接近法），④ method of seeking a maximum（仮説演繹的接近法））がある^{7,8)}。① Gestaltは臨床医による直感的なパターン認識であり，医師個人々人による個人差が激しく，質の管理が困難である。また，確証バイアスのため誤診に気づきにくい。② exhaustive enquiryは患者に関する情報を網羅的に集積し，総当たりに診断のあらゆる可能性を検討していくもので，効率が悪く，医療資源や患

者に対する負担が大きい。③ multiple branching processはアルゴリズム診断と呼ばれるもので，器官系や症候ごとに診断を煮詰めていく方法である。これもまた，範疇ごとに決定論的に診断していくため効率が悪く，アルゴリズムを順に追っていく必要があるので臨床場面での融通が利かない。④ method of seeking a maximumは，これは可能性の高い診断に向かって徐々に歩を進めていく方法で，演繹的で病態生理学的な仮説－検証のプロセスといえよう。ただし，この方法は蓋然性の高い診断名が多数存在するときに，診断の方向性が決め難い。

筆者らが開発した3つのAI診断システムは，臨床推論の方法論としては，② exhaustive enquiryと③ multiple branching processを折衷したものと考えられ，臨床医が通常用いる診断アプローチである① Gestaltの欠点である確証バイアスを補い，④ method of seeking a maximumの陥りがちな診断の迷宮から臨床医に進むべき代替的な方向性を指し示す優れたツールとなりうる。しかしながら，② exhaustive enquiryによる疾患領域の焦点の定まりにくさや③ multiple branching processの持つ人体機械論的な隘路からは自由ではない。

症例1の出題者はAIが得意とするキーワード／データベース検索では診断が困難であると思われる希少疾患で，かつ非典型的な症状を呈する疾患を選定した。一つの疾患治療中に別の症候が出現したという経過から医原性疾患を想起するという① Gestaltが働けばヒトに軍配があがることを期待した症例であるが，3か月以上前に開始した薬剤が原因であったことからヒトでも診断は容易ではない症例であり，会場でイソニアジド中毒を想起できたのは少数であった。症例提示には質疑応答も含め50分のプレゼンテーションで詳細な情報を与えたが，豊富な臨床情報のために逆に全体像を見誤った可能性もあり，これは日常診療においても起こりうる事である。今回，薬物副作用を診断の上位に位置付けたAIによる診断は臨床的に難解な症例において医師の助けとなることが期待された。

症例2については，① Gestalt，③ multiple branching process，④ method of seeking a maximumのいずれでも，正診にたどり着くことができるであろう。ただし，① Gestaltで正診を思いつのは，よほど呼吸器科での診療経験が豊富であるか，偶然に類似の症例を過去に経験しているか，といった，相当に困難な臨床推論経路である。③ multiple branching processは最も無理のないところで，症状が呼吸器科領域であることに鑑み，④ method of seeking a maximumと合わせて診断を追求していけば正診にたどり着く。

症例3は非特異的な臓器症状や全身症状が多発している

ことから、① Gestalt、③ multiple branching process、④ method of seeking a maximumでは正診を思いつきにくい。② exhaustive enquiry がもっとも有効な診断戦略となるが、限られた医療資源と時間的制約の中では、②は実践しがたい。また、この症例では発熱、消化器症状、急性心不全があり、中枢神経症状がみあたらないことから、ここから甲状腺クリーゼを疑うことは困難である。これを臨床現場で迷いなく診断可能であったことは、この症例を提示した医師の① Gestalt が並外れていることを示唆する。

以上の結果を踏まえて機械学習を用いた AI 診断システムの今後の可能性について考える。

T-DR においてヒトが疾患固有のデータ群（症状・所見データ）を分析しシステムの係数などを設定していた³⁾のに対し、機械学習の利点の 1 つはデータさえ与えれば AI が自分で学習するため症例の学習、追加学習が容易なことである。つまり I-DR・K-DR では出力結果が正診と遠くかけ離れていてもその症例を追加で学習させ（教師つきマシン・ラーニング）ておけば、AI 診断システムへの入力データが同じならば、次回からは上位に提示できるようになる。本症例では症例 2、症例 3 についてはその場で学習させることで、正診となる疾患を 1 位に挙げることができるようになった。機械学習を用いた AI 診断システムは学習を重ねることで徐々に出力結果の精度が上がってくる。このことは表 6 からもうかがわれる。

AI 診断システムの出力精度を高めるためには疾患固有のデータ群を集めるのが最も重要である。しかしこのなかに質の低い症例（間違っ て診断された症例）やデータ（直接的には当該疾患とは無関係な症状・所見）が交じると正確性は損なわれる。つまり質の高いデータを数多く集める必要があるがこれは現実にはかなり難しい。また集めたデータはテキストデータであるため、現在は AI 診断システムが学習できるように必要なデータを選択し、加工（コード化）するのをヒトが手作業で行っている。そのため学習させる症例とその症状・所見データを準備するのに多くの時間を費やしている。この点現時点で開発が進んでいる画像診断システムは画像とその答えだけを集めればよいのでデータ収集が容易である。実際 Web 上には画像診断システムに学習させるためのデータセットも多く公開されている。

AI 診断システムにテキストデータとして数多くの症例とそれぞれの固有のデータ群を集めるために、いわゆる電子カルテから情報を取り出すという方法が考えられるが、その現状に鑑みれば質が保証されない。一方できるだけ正確に固有データ群の収集を求めた場合、データの構造やフ

ォーマット設定からはじまり、1 疾患あたりにどれだけの固有データ群が必要なのかさえ定かではなく、質は保証されるとしても解決すべき問題が山積している。

ヒトと AI 診断システムの関係について述べる。筆者らの AI 診断システムは可能性のある疾患を可能性の高い順に提示することを目的として作成された。これは鑑別疾患として思いつくべき疾患を思い出せない・思いつかないというヒト（医師）の診断エラーを減らすことに大きな意義があると考えられる⁹⁾。

しかし筆者らの AI 診断システムはただ過去の症例をもとに入力された症状に対して可能性の高い疾患を挙げるだけである。提示された疾患情報をもとに考え、最終的診断を下すのは熟練したヒトであろう。またシステムに入力するデータを決定するのも医師でないと難しいと考える。筆者らは「ヒト vs AI」という構図ではなく、ヒトが患者からの確に所見を取り、その上で AI 診断システムに大量の情報を処理させつつその支援のもとでヒトが分析結果を吟味して最終的な決断を下し治療を行うというかたちで協力していくことが医療の質の向上につながると考えている。

本論文内容に関連する著者の利益相反はない。

文 献

- 1) Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, et al.: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* (2017) 542, 115-118.
- 2) Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, et al.: Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol* (2018) 29, 1836-1842.
- 3) 鳥越恵治郎: コンピュータ診断支援ツール. *日本医事新報* (2003) 4120, 24-32.
- 4) 鳥越恵治郎: コンピュータ診断支援ツール (病名思い出しから病名絞込みへ). *日本医事新報* (2006) 4310, 73-75.
- 5) Torigoe K, Tokuda Y: Potential Usefulness of Diagnostic Reminder as Web-based Clinical Decision Support System. *JHS* (2016) 4, 297-303.
- 6) 株式会社アイロム: 新病名思い出しツール. <http://mith.akira.ne.jp/iromtest/diagnosis2/> (2018年 6月閲覧)
- 7) Murphy EA: *The Logic of Medicine*, 2nd ed, Johns Hopkins University Press, Baltimore (1997) pp305-307.
- 8) 古川俊之: 医師の診断のアルゴリズム. *数理科学* (1983) 241, 5-12.
- 9) Torigoe K, Tokuda Y, Matsumoto K, Yasuda H: Evaluation of a Web-based Diagnosis Reminder System for Difficult Diagnostic Reasoning. *JHS* (2017) 5, 107-109.