



# ITU-TフォーカスグループFG-AI4H

慶應義塾大学大学院 政策・メディア研究科 特任教授

かわもり まさひと  
川森 雅仁



## 1. はじめに

FG-AI4H (Focus Group on AI for Health) は、ITU-TがWHOとの協力の下に2018年7月に設立した時限のフォーカスグループである。2018年5月にITU本部で開催されたAI For GoodワークショップにおいてAI for Healthのフォーカスグループが話題になった。これを受け、2018年7月のSG16会合に併設してAI for Healthのワークショップが開催された。またSG16会合においてWHOと共同でITUが主催するFG-AI4Hの発足が承認された。2018年9月に、最初のFG-AI4HがWHOの主催でジュネーブにおいて開催された。その後、第2回会合が2018年11月14日～16日にニューヨークのコロンビア大学で、第3回会合が2019年1月22日～25日にローザンヌのローザンヌ・スイス連邦工科大学 (EPFL) で開催された。第4回会合は、2019年4月2日～5日で、中国情報通信研究院のホストで上海で開催された。

## 2. フォーカスグループの体制

現在、FG-AI4Hは、以下の共同体制で議論を進めている。中心的な役割はThomas Wiegand氏 (Fraunhofer HHI, Germany) とMarcel Salathé氏 (EPFL, Switzerland) が担っている。残りの共同(副)議長はほぼ毎回増減があるので、名誉議長的存在と言える。

議長: Thomas Wiegand氏 (Fraunhofer HHI, Germany)

副議長:

- Marcel Salathé氏 (EPFL, Switzerland)
- Ramesh Kishnamurthy氏 (WHO)
- Sameer Pujari氏 (WHO)
- Shan Xu氏 (CAICT, China)
- Stephen Ibaraki氏 (ACM and REDDS Capital, USA)
- Naomi Lee氏 (The Lancet, UK)

Thomas Wiegand氏は、ITU-T SG16のQ6のラポーターでありH.264、H.265の勧告化をMPEGと推進した経験を持つ。その経験を活かして、AI for Healthのリーダーシップを取っている。Marcel Salathé氏は、医療と技術の両面に明るい研究者で、今回のFGの技術上の中心的存在である。

FGは、ITUメンバー以外も参加できる、オープンなグループのため、開催地域によって、参加者の大部分が入れ替わ

る。ほぼ恒常的に参加しているのは、共同議長を務めるFraunhofer HHI (ドイツ) とEPFLのスタッフや学生が中心となっている。WHOで開催された第1回会合には、20か国から約100名が参加した。この中には遠隔参加者も含まれる。しかし、第2回以降は、すこし少ない人数になっている。

## 3. 健康医療とAIの例

第1回会合は、特にWHO主催であったこともあり、WHOの関与が大きかった。これはFG-AI4Hに対するWHOの期待の大きさも示すと言える。AIの応用が期待される医療健康領域の応用例として、世界経済フォーラムの記事の例を引用する形で、紹介された。以下に第1回会合で紹介された医療健康領域のAI応用の例を示す。

### 皮膚がんの検出

2018年5月22日に発表された医学論文によると、皮膚がんの画像とそれに対応する診断を用いて機械学習AI診断システムは、95%の皮膚がんの検出率を達成したのに対し、実験に参加した58人の皮膚科医は、87%の検出率しか上げられなかった、という。このように皮膚科の世界では、現在AIによる診断システムの開発が進んでおり、FG-AI4Hにも皮膚科医が何人か参加していた。

### 眼病診断

糖尿病性網膜症 (DR) と加齢性黄斑変性症 (AMD) は、英国で62万5000人以上、そして世界中で1億人以上が罹患しているもので、失明の主な原因となっている。

Google DeepMindは、この2つの症状の診断に取り組むために、ロンドンのMoorfields Eye Hospitalと提携し、2018年7月に、その最初の成果をNature Medicineに発表した。

この論文によると、それまで未解決であった実際の臨床環境で、三次元診断画像による診断において、専門臨床医と同程度の能力を達成するという課題に対する解決の糸口を作った。ここでは、患者からの臨床的に不均一な三次元OCT画像に、新たに開発されたディープラーニングアルゴリズムを適用することにより、14,884枚の画像で機械学習させた後に、失明につながる一連の網膜疾患について専

専門家と同程度かそれ以上の診断を行うことができたという。このように眼科分野でのAIの応用は期待されており、FG-AI4Hでも、眼科への応用が度々提案された。

## 医療ロボット

中国深センのiFlytekが開発したロボットは、実際の医師試験で456点という高得点を達成した（合格するのに必要なのは360点）。これを成し遂げるために、このロボットは医者になるために必要な種類の推論を開発するための何十もの医療教科書、200万件の医療記録及び40万件の医学論文の内容を学習したとされる。また、単に記憶するだけでなく、文間の関係を推論する能力も備えているという。ちなみに、iFlytekは、翻訳ソフトや音声認識エンジンの製品化でも有名で、AI全般に力を入れた会社である。

## FastMRI

Facebookがニューヨーク州立大学と共同研究を行っている、FastMRIについての発表があった。このプロジェクトは、AIを用いて、MRIのスキャンに要する時間を10分の1にしようというものである。まだ、研究途上のものであるため、具体的にどう早くするか、ということは、述べられなかった。会場にいた専門家からは、MRIの理論的、技術的背景が非常に複雑なものである故、なかなか思うような結果は出ないのではないか、というような意見があった。

## AI内視鏡

インターネットビジネスで有名なTencentが、内視鏡映像をAIで分析し、網膜のガンを発見する、というシステムを紹介していた。これは、上記のGoogleの模倣のような感じだが、iFlytekの例と並んで、中国でのAIへの関心の高さと普及の速さを示していると言えよう。

## 4. 健康医療におけるAI技術の標準指標の必要性和スコープ

前述のように、健康医療分野で多くの応用が期待されるAIだが、そのアルゴリズムの複雑性のため、AI依存の技術の良し悪し、長所短所を判断するのが難しくなっている。このことが、AI技術の信頼性に対する障壁となっている側面があり、WHOのような機関が、AIの使用を健康医療分野で推奨する上で問題となっている。その問題を解決するために、現在のAI技術に関する指標の標準化は、健康医療分野にとって有益なものと考えられる。

このような背景から、FG-AI4Hは以下のような標準的項目を規定するための枠組みを提案するとしている。

- ・AI依存の技術に関する指標
- ・責任の所在と信頼のためのガイドライン
- ・世界的なレベルでAIを健康問題に適用するためのオープンなデータとプロセス

しかしながら、FG-AI4Hは、AIアルゴリズムそのものは標準化しない。またフォーカスグループであるため、ITU勧告を作成することはしない。勧告化作業は、FG-AI4Hの親SGであるSG16に委ねられる予定である。

## 5. FG-AI4Hの具体的標準化手法

FG-AI4Hの標準化手法は、これまでの一般的なギャップアナリシスから始まるものとは異なる。以下にその概要を述べる。

### 5.1 ベンチマーク・プラットフォーム

現在、一般にAIと呼ばれる技術は、その多くがニューラルネットワークによる機械学習、特にDeep Learningという技術に拠っている。ここで言うニューラルネットワークとは、生体のニューロンの動作を極めて簡易に表現した、入力層と出力層、そしてその間にある中間層からなるノード（人工ニューロン）の集合からなるネットワークである。その内、中間層が4つ以上存在するようなものを多層ニューラルネットワーク、さらにそれより多い中間層を持つものを深層ニューラルネットワークといい、このようなネットワークに典型的に代表されるような機械学習法をDeep Learningと呼ぶ。現在のDeep Learningアルゴリズム自体は、例えば医療や健康のような領域に特化した専門知識（Domain Knowledge）を特に必要としないとされる。つまり、良いデータ・セットを持っていれば、誰でもDeep Learningネットワークを訓練しAIシステムを作ることができると考えられている。このことには以下のような長所と短所が存在する。

- ・長所：非常にダイナミックな分野：外部からの新規参入が楽
- ・短所：AIモデルの（良し悪しを決める）比較法がない。

このことは、国、政府、あるいは規制側にとっては、医療領域でAIを使用する場合に(a)何らかのベンチマーク（基準）と(b)関係者の合意に基づく基準設定として後述の2つを標準化することが必要になるとされる。

FG-AI4Hでは、これらの要求条件に基づいて、AI、特に機械学習を医療、健康保健の領域で使用するための、



技術的枠組みを標準的に提供することを目指す。また、WHOが参加することにより、技術的要求条件に加え、専門的な観点からの要求条件も、シナリオという形で記述する。

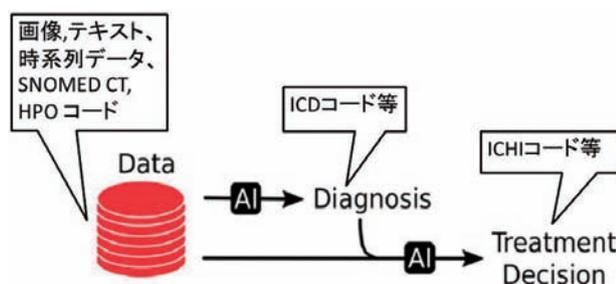
以上のことから、FG-AI4Hでは、以下のような項目について標準化を進める方針である。

- ・入力データセットの標準化
  - ・(医師により) 認定された各患者の診断結果の標準化
- 上記の標準化されたデータセットに基づき、以下のプロセスを行う：

- ・データセットを学習用データ(公開)とテスト用データ(非公開)に分離
- ・比較のための尺度基準を策定(時間やコストも基準に入れる)
- ・様々なAIアルゴリズムをこの基準に従って競争させる

最初のデータセットを学習用とテスト用の2つのセットに分離するのは、機械学習で一般に行われていることであり、それをFGの作業として行う。公開された学習データにより、FGの参加メンバーがAIを学習できる。ここで重要なことは、FGの目的は、AIアルゴリズムそのものを標準化するのではなく、入力データと出力データの間に関する基準(ベンチマーク)を標準化することである、ということである。

図1にこの関係を示す。



■ 図1. 入力データと治療判断の関係例 (FG-AI4H)

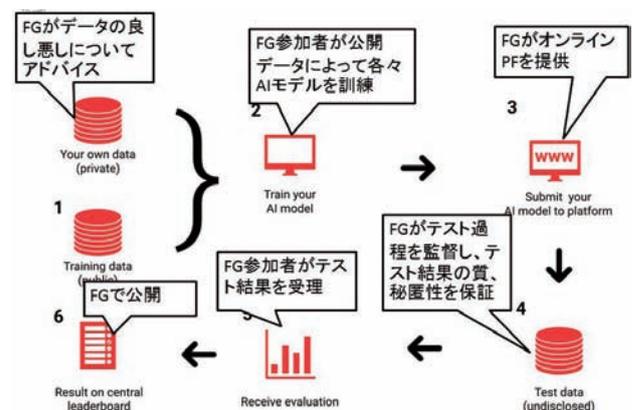
図中の略号の意味は以下のとおり：

- ・SNOMED CT (Systematized Nomenclature of Medicine—Clinical Terms) 英国ロンドンにある非営利標準開発団体SNOMED Internationalによって維持されている、臨床文書に使用されるコード、用語、同義語及び定義等の医学用語を体系化したコンピュータ処理可能な辞書。
- ・HPO (Human Phenotype Ontology)：人間個々の表現型異常を表す8000以上の用語からなるオントロジー。HPO Consortiumによって管理されている。
- ・ICD (International Statistical Classification of

Diseases and Related Health Problems)：WHOによる国際疾病統計分類。

- ・ICHI (International Classification of Health Interventions)：WHOが開発している介入手順に関する分類。

図2は、上記のベンチマークの具体的な流れを記述したものである。



■ 図2. ベンチマーク枠組みの概要

この図から分かるように、FG-AI4Hは、参加者のデータを集積するとともに、その一部を学習データとしてオンラインで公開提供し、参加者が機械学習をすることを促すとともに、クローズドな環境でテストし、その結果を公表する、というサイクルを繰り返すことを期待している。このような枠組みにより、FGの言うベンチマークが形成され、様々なAIコミュニティがこの枠組みとベンチマークを採用することにより、最初に挙げた、AIの信用を得ることができるようになる、という方向性になる。

またベンチマーキングに用いる、AIのタイプについては、WHOから次ページの表のような分類の提案があった。これは出力文書A-104 “Draft thematic classification scheme”に含められている。ただし、これは、WHOが試みた分類例であり、まだ最終形ではない。また、その分類の妥当性についても、専門家を交えた議論がまだ十分されておらず、このままで採用されるとは思えない。

## 5.2 Task Group (TG) の構成

FG-AI4Hでは、第3回のローザンヌ会合で、以下のTGに分かれて、議論が推進されることになった。

1. 皮膚科 (Dermatology ; TG-Derm)
2. 高齢者転倒問題 (Falls among the elderly ; TG-Falls)
3. 病理組織学 (Histopathology ; TG-Histo)

■表. WHOによるAIとデータの分類

Level-2 Thematic Classification (Artificial Intelligence)
<i>AI-benchmarking class type</i>
1. Machine Learning
1.1. Classification
1.2. Regression
1.3. Clustering
1.4. Recommendation systems
1.5. Matching
1.6. Sequential data models
1.7. Anomaly detection
2. Optimization
2.1. Automated planning & scheduling
2.2. Evolutionary algorithms
3. Knowledge representation and reasoning
3.1. Default reasoning
3.2. Common-sense knowledge
3.3. Ontological engineering
3.4. Sub-symbolic reasoning
4. Artificial Intelligence
4.1. Generative models
4.2. Autonomous systems
4.3. Distributed systems
5. Perception
5.1. Visual recognition (photo/video)
5.2. Natural Language Processing (text/voice)
6. Affective computing
6.1. Sentiment analysis
6.2. Virtual assistants
6.3. Social agents
7. Motion and manipulation (robotics)
7.1. Robotic arms
7.2. Industrial robots
7.3. Motion planning
8. General intelligence

4. 神経認知疾患 (Neuro-cognitive diseases ; TG-Cogni)
5. 眼科 (Ophthalmology ; TG-Ophthalmolmo)
6. 精神科 (Psychiatry ; TG-Psy)
7. 蛇識別 Snakebite and snake identification (TG-Snake)
8. 症状評価 (Symptom assessment ; TG-Symptom)
9. 結核 (Tuberculosis ; TG-TB)
10. 心血管疾患リスク予測 (Cardiovascular disease risk prediction ; TG-Cardio)
11. 脳画像の分析による自閉症の分類 (Classifying autism through analysis of brain imagery ; TG-Autism)

2019年4月の第4回上海会合時点では、TGの中にはデータを募集しているものもあり、野心的ではあるが、2年とい

う時限で行われるFG活動内で、11の分野で十分な結果が出るかどうかは、未知数である。

### 5.3 FG-AI4Hの今後の予定

第4回会合に引き続き以下の会議が現在のところ予定されている。

- ・第5回会合：2019年5月29日～6月1日（ジュネーブ）
- ・第6回会合：2019年9月2日～5日（ザンジバル）
- ・第7回会合：2019年11月（ニューデリー）
- ・第8回会合：2020年1月（ブラジリア）

## 6. おわりに

健康、医療、福祉の分野では、そもそも初期のAI技術の最初の適用分野が医療における診断補助用のエキスパートシステムであったことから分かるように、AIの有効な適用領域とされてきている。本稿で見たように、近年、さらに様々なアプローチで医療にAIを適用するという例が報告されてきている。しかしながら、これらの様々なAI応用のアプローチには、標準化という視点が、往々にして欠けており、その結果、AI技術の適用方法やその結果についての疑念を招くことになりかねない事態も起きている。それ故、異なったアプローチのAIの健康、医療、福祉分野への適用を客観的に評価できる標準が望まれている。このようなギャップを埋めようと、WHOと協力して立ち上げられたのが、FG-AI4Hである。

しかしながら、FG-AI4Hには以下のような課題があると思われる。

- データセットの収集が急務であるが、権利関係やプライバシー等、解決すべき課題が多い。
- WHOが期待している広範にわたるAIの応用に対して、FGは、時限機関であるため、限られた時間で広範な領域のデータを集め、ベンチマークの構築と標準化を同時並行に行う必要がある。
- FGがWebプラットフォームを構築することになっているが、具体的な方法が、まだ不明。（共同議長の所属するFraunhoferとEPFLが用意することになると想定されるが、それが標準と呼べるか。）
- 現在すでに乱立気味のAI関係の団体やイベントとの関係をどうするか。

以上のような、課題は、存在するが、AIに対する大きな期待と関心のもとに、FG-AI4Hは今後も活発な活動を行うと期待される。それ故、今後の動向を十分に知る必要がある。