

感染症調査での AI の進歩(総説) NEJM, April 27, 2023

「僻地で世界最先端」西伊豆健育会病院早朝カンファ 2023 年 5 月

Advances in Artificial Intelligence for Infectious-Disease Surveillance
(Review Article)

著者

- John S. Brownstein, Ph.D., Benjamin Rader, M.P.H., Christina M. Astley, M.D., Sc.D.,
The Computational Epidemiology Laboratory, Boston Children's Hospital,
Harvard Medical School of Public Health, Massachusetts, U.S.A.
- Huaiyu Tian, Ph.D.
Center for Global Change and Public Health, Beijing Normal University, Beijing

NEJM, April 27, 2023 に「感染症サーベイランス(調査)での AI の進歩」の総説がありました。
1 カ月前、NEJM, March 30, 2023 の総説は大興奮の「臨床医学での AI と機械学習」でしたが、
今回はその実際の具体例が示されています。

NEJM, April 27, 2023「感染症サーベイランスでの AI の進歩」最重要点は下記 9 点です。

- ① 感染症サーベイランスに AI は極めて有効。中国の COVID 初発数日で肺炎クラスターを検出した。
- ② 症状検出のみでなく病理学的証拠が必要。抗菌薬阻止円をスマホ撮影、アプリで判定報告。
- ③ AI の遺伝子検索と電カル関連付けで発症ルート判明。技師が MRSA 伝播、胃カメラで緑膿菌伝播。
- ④ 中国は QR コードのスマホ利用で位置追跡、ギリシャは強化学習 AI で入国者の PCR 対象決定。
- ⑤ スマートウォッチ、スマートリングで心拍から患者発見。画像でマスク装着率、動き、社会的距離わかる。
- ⑥ AI はゴミを入れたらゴミしか出てこない (Garbage in, garbage out)。
- ⑦ スマホ情報のプライバシー保護に連合学習 (federated learning) 有効。
- ⑧ GPT-4 は感染症サーベイランスに極めて有望。
- ⑨ この総説の Chatbot GPT-4 によるまとめ、グーグル翻訳の悪文。

2023 年 3 月 16 日、マイクロソフト社は Microsoft 365 Copilot の発売を発表しました。
従来の Word、Excel、PowerPoint、Outlook、Teams に AI を組み込むと言うのです。
下記の仰天の動画をご覧ください。たとえば「ある主題で 10 枚のパワーポイントの動画を作ってくれ」と頼めば数秒で作成、「データから利益のある方法を教えてくれ」と言えばあっという間に示してくれるのです。

[Introducing Microsoft 365 Copilot - your copilot for work - The Official Microsoft Blog](https://www.microsoft.com/ja-jp/copilot)

Excel に AI を組み込んだらどういうことになるのか？ あと半年ほどで事務や会計業務の大量失業時代が到来すると言うことでしょうか。またプログラムを監視する人員は必要であつてもプログラマー自体は不要となります。市役所の人員も大幅な削減が可能でしょう。営々として築いてきた仕事が一瞬で無価値となる凄まじい産業革命が現実のものとなります。

みずほ銀行では 2023 年 5 月現在、他の銀行への 3 万円以上の振り込み手数料が ATM のカード決済なら 330 円ですが窓口で人手を介するとなんと 880 円です。
窓口業務を縮小し人員削減を行うということでしょうか。

今後、更に量子コンピューターが AI と結びついたらいったいどんなことになるのか空恐ろしくなります。コンピューターは 0 か 1 の 2 進法ですが電子軌道は 0 と 1 の両方の値を同時に取り得ます(量子重ね合わせ)から超並列的处理が可能となります。

医療も影響を受けないはずはありません。画像分析(皮膚科、病理、放射線科)は AI が最も得意とするところですから画像診断は AI にとって代わるでしょう。下記の NEJM 総説をご覧ください。近未来の皮膚科診療が描かれています。皮膚疾患はスマホで診断可能となるのです。

[conference_2019_08.pdf \(nishiizu.gr.jp\)](https://www.nejm.org/doi/full/10.1056/NEJMe20190801)

(医学での機械学習、Review Article、NEJM, April 4, 2019 西伊豆早朝カンファ)

1. 感染症サーベイランスに AI は極めて有効。中国の COVID 初発数日で肺炎クラスターを検出した。

この 2023 年春から伊豆半島でも急速に来訪客が増え、観光が力強く復活してきました。
過去 3 年間、COVID のパンデミックでは、保健所の方々の激務に本当に頭が下がります。
当院でもクラスターを起こし献身的にサポートして頂きました。
特に感染者が発見されたあとの保健所の接触者追跡はあまりに労働集約的(labor intensive)な作業で膨大な労力、人手を要しました。

Web 上にはニュース、報道会見(press releases)、専門家会議などの情報があふれており、その中にウイルス感染初発情報もあります。しかし情報量があまりに膨大で感染症専門医であつても初期アラート(警告)を発することは不可能でした。

「Health Map」は過去 10 年以上にわたり使われてきたインターネットに基づく感染症サーベイランスシステムです。メキシコでインフルエンザ A(H1N1) の出現を早期に検知し、2019 年米国の vaping induced pneumonia(電子タバコによる肺炎多発。電子タバコは COPD の発症リスクでもあります)も追跡できました。また Health Map は 2019 年 12 月 30 日、中国での COVID19 患者初発後、早くも数日で「原因不明の肺炎クラスター発生」を感知できたのです。

Health Map は自然言語処理(人間の言語のテキストデータを機械で処理、内容を抽出)により 9 か国語以上の言語で Web 上を横断的に感染症発症をリアルタイムで検索します。
その際、病原菌と地理の辞書を検索に使用します。そのアルゴリズム(計算)では雑音を無視し疾患関連のテキストから病原菌の名前、発症数などを解析(parse)しベイズ推論(Bayesian inference)を行います。

例えばシアトルの報道で下記の文があると、検索エンジンはこの中から病原体名 coronavirus、地名 Seattle、発生数 6 人のデータを抽出し、合計数 14 人は無視します。同様の報告を集めて結合しワシントン州全体の地域別疾患発生をカウントするのです。

「New cases of novel coronavirus in Seattle area spark concern among local health officials. Six individuals diagnosed, which brings Washington's total to fourteen so far.」

機械学習のベイズ推論とはまず予測する確率(probability)を「70%位じゃね？」とエイヤと適当に決め(事前確率の設定)、抽出したデータから最初のいい加減なデータを補正し続け(事前確率補正)、そして精度の高い推論に近づける(事後確率の算出)手法です。データがたくさん集まれば正規分布曲線を描くことができ確率(probability)と尤度(likelihood)がわかります。例えば偏差値 70 点であればそれより上の正規分布の下の面積が 2.28%であり、これが確率、70 点の時の正規分布の Y 軸が 1000 人中 22.8 人でありこれが尤度です。

AI と機械学習による方法は、従来の労働集約的(labor-intensive)な人力によるサーベイランスに比しはるかに有用(key advantage)です。

Covid 19 の危機で AI、機械学習は、COVID19 とインフルエンザを区別、発見する初期アラートシステム(early warning)や発症地発見、疫学的追跡、予測、医療資源配分等に極めて効果的であることがわかりました。

まとめますと感染症サーベイランスに AI、機械学習は極めて有効でした。中国の COVID 初発数日で肺炎クラスターを検出することができたのです。

2. 症状検出のみでなく病理学的証拠が必要。抗菌薬阻止円をスマホ撮影、判定報告。

COVID-19 に似た症状は様々な疾患で起こり得ます。単に症状のみに頼ると、カナダで実際あったように「COVID 様症状」が enterovirus や rhinovirus に依るのにも関わらず COVID の波と間違えます。アウトブレイクの病理的特徴(遺伝子検査など)を定義し症状と関連付けることにより公衆衛生機関はインフルエンザに対する oseltamivir (タミフル)の確保準備など迅速な対応ができます。

The Kirby-Bauer disk-diffusion test は細菌培地に抗菌薬の disk を置きその周辺に現れる阻止円の直径を計測してその抗菌薬の耐性の有無を判断する方法です。しかしその計測は多分に主観的(user-dependent)です。国境なき医師団(Médecins sans Frontières)はこの判断に AI によるアプリを利用しました。

培地の阻止円をスマホで撮影しアプリが画像処理、この直径を測定し感受性、非感受性を判定し抗菌薬名とその結果を WHO のサーベイランス機関(the Global Antimicrobial Resistance Surveillance system of WHO)に送信するのです。

これにより世界的規模で抗菌薬抵抗性を追跡(track)できます。

まとめますと AI によるサーベイランス(調査)は症状検出のみでなく病理学的証拠との関連付けが必要です。国境なき医師団は抗菌薬阻止円をスマホ撮影しアプリで判定、WHO に報告しました。

3. AI の遺伝子検索と電カル関連付けで発症ルート判明。技師が MRSA 伝播、胃カメラで緑膿菌伝播。

アウトブレイクを一旦確認したら次のステップは患者の追跡(trace)、感染経路の遮断です。従来クラスターの時間、空間的追跡は人力で行われてきましたが極めて労働集約的(labor-intensive)であり大規模なカルテ追跡(chart review)、環境のサンプリング、インタビューが必要となります。

この数年でピッツバーグ大学ではアウトブレイク源確認のため whole-genome surveillance (全遺伝子検査)に the Enhanced Detection System for Healthcare-Associated Transmission (EDS-HAT)を用いました。これは AI による全遺伝子検索と電子カルテ検索を関連付けたものです。感染した患者の電子カルテだけでなく他の病院の似たような患者の電子カルテも検索します。EDS-HAT により 9 か所の病院での疾患発生を 40%近くまで防ぐことができたのです。

また従来だったら到底関連付けられぬような例も電子カルテから見つけることができました。院内の異なる病棟で MRSA(メチシリン耐性ブドウ球菌)患者が発生、発症は 8 日間離れていました。しかし同一の検査技師がベッドサイドで心電図を取ったことがわかり、これにより伝播したと考えられました。

また同じ遺伝子型の緑膿菌患者が 6 名、別々の病棟で発生しましたが時間と空間が離れていたため当初、無関係と思われました。しかし機械学習のアルゴリズム(計算)により同一の胃カメラが感染経路とわかったのです。

ただ小生、思ったのは現在、国内では各病院の電子カルテに互換性がありません。同一病院内なら検索ができますが、他の病院の電子カルテは検索できません。例えば hepatoma は肝臓、肝がん、HCC と表記は様々です。このままでは AI は同じものと認識できません。Hepatoma の「概念」に数字の識別子を付け、その概念の「表記」(description)として複数の用語(肝臓、肝がん、HCC、hepatoma)を関連付けることにより用語は同義語となります。

同じ概念に複数の用語がある時、複数病院間のデータ比較はできません。ソフトウェアシステムでは用語の明言(assertion)が必要なのです。

コンピューター言語は最終的には0と1でありその中間はありません。

あいまいな言い方は許されないのです。

異なる分野のフォーマット、構造を統合しコンピューター処理可能(computable)とし分類可能(classifiable)とする必要があります。これを行うプラットフォームを情報工学ではオントロジーと言います。共通のオントロジーにより基礎科学データ、遺伝情報、電子カルテ間のやり取りが初めて可能となるのです。米国では既にオントロジーが統一されているということなのでしょうか？

オン(ov)とはギリシャ語の「存在」のことで ontology はもともと形而上学のことです。
「神は存在するのか?」「時間の始まり、終わりはあるのか?」「最小単位はあるのか?」などの哲学をすることが ontology (形而上学)でした。

下記の NEJM の精密医療の総説では共通の ontology を作ることが精密医療の一丁目一番地だとしています。

[conference-30_18.pdf \(nishizu.gr.jp\)](https://www.nishizu.gr.jp/conference-30_18.pdf)

Precision Medicine (精密医療)(総説) NEMM, Oct.11, 2018 西伊豆早朝カンファ

まとめますと AI の遺伝子検索と電子カルテの関連付けで発症ルートが判明します。
同一検査技師が MRSA を伝播、胃カメラで緑膿菌伝播された事例が AI によってわかりました。

4. 中国は QR コードのスマホ利用で位置追跡、ギリシャは強化学習 AI で入国者の PCR 対象決定。

患者のリスク評価にも AI は有用です。

広範なパンデミックでは単一予防策による根絶は無理であり、ワクチン接種、接触歴追跡 (contact tracing)、移動制限、マスク装着等様々な方策が必要です。

中国では QR(quick response)codes がスマートフォン (Alipay、WeChat) で使われリアルタイムの位置追跡と AI による chatbot で健康関連の質問に答えました。

QR codes って quick response だったんですね。初めて知りました。

ギリシャ政府は国境での旅行者に AI アルゴリズムの Eva を用いて COVID の PCR を行う入国者を決めました。

即ち死亡率、陽性率の多い諸国の者を測定するような一般疫学的測定基準(metrics)でなく出身国、年齢、性、時期の変数を入力(features:特徴量と言います)して強化学習させました。

強化学習(reinforcement learning)とは試行錯誤のやりとりを重ねてタスクを実行、タスクの報酬を最大化するような意思決定を行ないます。

旅行者の全数検査をしたのではなく、この出力(labels)に従って PCR を行い医療資源を無駄にしなくて済んだのです。機械もまたこの結果から学習するのです。

機械学習はルールに従ってプログラムするのでなく機械が教材(examples)から学習していくのです。ある課題に対し教材(features)が入力され回答の出力(labels, 答え、目的変数)を与えてトレーニングします。これによりギリシャでは通常の疫学的方法より 1.25 倍から 1.45 倍の感染者を検出できました。

まとめますと AI は患者のリスク評価にも使われました。中国は QR コードのスマホ利用で患者の位置追跡、ギリシャは強化学習 AI で入国者の PCR 対象を決定したのです。

5. スマートウォッチ、スマートリングで心拍から患者発見。画像でマスク装着率、動き、社会的距離わかる。

更なる応用として、スマートウォッチ(手首に着ける)やスマートリング(指輪のように指に着ける wearable device)の情報から発症前から心拍数が高いなどの信号を検出できアウトブレイク切迫がわかります。またアウトブレイクの中心地(hotspots)を確定できます。待合室の患者の拾い出しも可能です。また画像認識アルゴリズムでマスク装着率、人の動き、社会的距離の取り方が判ります。

6. AI はゴミを入れたらゴミしか出てこない(Garbage in, garbage out)。

どんなに優れた AI 感染症サーベイランスシステムであっても入力データが偏っていた場合、その予測は信用できません。ゴミを入れたらゴミしか出てこない(Garbage in, garbage out)のです。従って AI は高品質の従来のサーベイランスシステムの補助とすべきでしょう。

また疾患追跡システムも遺伝子検査と連動していなければ似たような症状を呈する疾患を区別できません。また COVID-19 は頻回に新種が出現するため AI の再カリブレーション(校正)が必要です。入力データが不十分だと価値はなく有害ですらあります。

またサーベイランスシステムの精度が上がっても、もしそのデータベースで選択された人々に偏りがあった場合、不完全なものとなります。実際、米国の COVID-19 死亡率は、黒人とヒスパニック系での死亡が 60%近くまで未報告だったため不完全な結果となりました。

また COVID-19 診断で機械に画像分類をトレーニングさせる際、コントロール群として小児 X 線のデータベースを用いました。ところが機械のアルゴリズムは COVID-19 を診断するのではなく単に小児と成人の X 線を区別しただけでした。Padua 大学では胸部 X 線をデータベースから消去しても、どのデータベースから結論を導いたか推論できたのです。

すなわち AI は入力データが正しく偏りのないことがなにより重要であり、ゴミを入れたらゴミしか出てこない(gabage in, garbage out)のです。

7. スマホ情報のプライバシー保護に連合学習(federated learning)有効。

サーベイランスは digital exhaust(車の排ガスのようにインターネットを使用した残りカスで、個人が SNS に上げるような情報を指す)や wearable device(アップルウォッチなど)の情報も利用します。ですから個人情報の保護が必要であり、その解決法としては連合学習(federated learning)があります。

機械学習はふつう大量のデータを一つのサーバーに集約するためかなりの負荷がかかります。連合学習ではすでに学習が完了したデータのみを 1 か所のサーバーに送信するためサーバーへの

負荷が少なく、また元データがその持ち主から離れず、学習の結果のみをサーバーへ送信するのでプライバシーの確保ができるというのです。

最近、スマートフォンからの健康情報収集に連合学習 (federated learning) が使われています。

まとめますとスマホ情報のプライバシー保護に連合学習 (federated learning) が有効です。

8. GPT-4 は感染症サーベイランスに極めて有望。

Chatbot GPT-4 は将来の感染症サーベイランスに極めて有望と思われます。これは膨大な非構造化データ(標準的な大量のファイル)を分析して労働集約的作業を軽減、隠れた感染症流行地を見いだせます。

9. この総説の Chatbot GPT-4 によるまとめ、グーグル翻訳の悪文。

小生、Chatbot GPT-4 に、この NEJM 総説の要約を頼んだところ以下のとおりでした。

「NEJM, April 27, 2023 感染症サーベイランスに用いる AI の進歩についての要約：
感染症の集団発生を同定し、追跡し、抑制するために、AI と機械学習ツールを使用する方法を検討した総説である。感染症サーベイランスにおける AI の様々な機能を紹介し、例えば早期警戒システム、ホットスポット検出、疫学的追跡と予測、資源配分などが挙げられる。
また、公衆衛生サーベイランスにおける AI と機械学習の限界と課題についても議論し、将来的な実施を改善するための考察を示す。」

ところで皆さまにお聞きしたいのは、「グーグル翻訳」を皆様どのように利用されているのでしょうか？ 実はこの総説の一部をグーグル翻訳したのが下記の文です。
悪文の最たるもので小生、ついに理解できませんでした。英語だとそれほど難解ではないのです。おそらく日本語の機械学習の際、「教師あり学習、supervised learning」がまずかったのではと思います。学習の手本を示す時、国語の先生とか文系の教師の関与が必要ではないでしょうか。

【この総説の一部のグーグル翻訳】

「データ量と品質

大量の低遅延データの可用性は、感染症サーベイランスの改善に大きな役割を果たしてきましたが、ギャップは残っており、脆弱性を見逃され続けています。「ビッグデータの傲慢さ」は、AI で訓練された最も正確な感染症監視システムでさえ、過剰適合(つまり、特定のデータに合わせて調整されすぎているため一般化できない予測)につながる可能性があり、高品質の従来の監視を置き換えるのではなく補完する必要があることを思い出させます。」

この英文は以下の通りです。

Data Volume and Quality

The availability of large quantities of low-latency data has played a large part in improving infectious-disease surveillance, but gaps remain, and vulnerabilities continue to go unnoticed. “Big data hubris” reminds us that even the most accurate AI-trained infectious-disease surveillance systems can lead to overfitting (i.e., predictions that are not generalizable because they are too tailored to specific data) and should complement rather than replace high-quality traditional surveillance.

それでは NEJM, April 27, 2023「感染症調査での AI の進歩」最重要点 9 の怒涛の反復です。

- ① 感染症サーベイランスに AI は極めて有効。中国の COVID 初発数日で肺炎クラスターを検出した。
- ② 症状検出のみでなく病理学的証拠が必要。抗菌薬阻止円をスマホ撮影、アプリで判定報告。
- ③ AI の遺伝子検索と電カル関連付けで発症ルート判明。技師が MRSA 伝播、胃カメラで緑膿菌伝播。
- ④ 中国は QR コードのスマホ利用で位置追跡、ギリシャは強化学習 AI で入国者の PCR 対象決定。
- ⑤ スマートウォッチ、スマートリングで心拍から患者発見。画像でマスク装着率、動き、社会的距離わかる。
- ⑥ AI はゴミを入れたらゴミしか出てこない (Garbage in, garbage out)。
- ⑦ スマホ情報のプライバシー保護に連合学習 (federated learning) 有効。
- ⑧ GPT-4 は感染症サーベイランスに極めて有望。
- ⑨ この総説の Chatbot GPT-4 によるまとめ、グーグル翻訳の悪文。