

人工知能を用いた、塩分摂取の健康効果の検証のための、
メタボリックドミノの破綻に起因する
生活習慣病と死亡リスクエンジンの開発と個別化治療の実現化

山田 朋英^{1, 2)}, 庄嶋 伸浩²⁾, Chia-lin Lee³⁾, 野間 久史⁴⁾, 山内 敏正²⁾

¹⁾キングスカレッジロンドン生命医科学部, ²⁾東京大学大学院医学系研究科糖尿病代謝内科,

³⁾台中軍人病院, ⁴⁾統計数理研究所

概要

背景・目的 メタボリックドミノは世界における重大死因に係る生活習慣病の共通経路であり、その進展要因とされている食生活や運動習慣が与える影響を明らかにする必要がある。本研究の目的は、人工知能による臨床エビデンスの統合と体系化を実現するため、各個人の生活情報や他のバイオマーカーなどの膨大な個人情報から、人工知能を用いて統合解析を進めることである。本研究では、1)ヘルスケアの個別化アプローチとして、特に塩分摂取を促す(減らす)ことで、将来の死亡率や種々の致命的疾病リスクを下げるができる個人(集団)を人工知能で特定する。すなわち、この個人(集団)を特定することで、適切な個別化された塩分摂取量の推奨による健康効果を具体的に提案できる。また、2)ヘルスケアの集団的アプローチとして、診療ガイドラインの礎であるシステムティックレビューの自動化を目標とし、塩分摂取の過去の臨床知見の統合を目的とする。

方法 1)のヘルスケアの個別化アプローチとして、コホートデータを用いたシミュレーション解析を行った。2)のヘルスケアの集団的アプローチとしては、すでに出版済みの塩分摂取に関する介入試験の系統的レビューについて検討を行った。ニューラルネットワークを介した人工知能は、文章を読んだ人が重要だと判断する暗黙の基準を学習(ディープラーニング)し、多数の文書からその基準に沿ったものを抽出できる。教師データを学習する際に、教師データに含まれる単語ごとに教師データとの関連性と単語同士のつながりに関して伝達情報量を通して数値化し、ニューラルネットワークを介し、特徴量として利用し、これを用いた。

結果 1)のヘルスケアの個別化アプローチとして、3万人規模のコホートにおいて、塩分の摂取を増加させる(あるいは減少させる)介入により、予後が改善する個人を推定した。シミュレーションの結果5年死亡に関するリスク低下の傾向が認められた。一方、2)のヘルスケアの集団的アプローチにおいては、ディープニューラルネットワークベースの機械学習により、系統的レビューのスクリーニング作業負荷が90%軽減されたことを報告した。

考察 1. データ収集とデータ蓄積, 2. データ前処理・学習・推論エンジン生成, 3. 開発・実行・モニタリング研究を進めていく。

1. 研究目的

メタボリックドミノは世界における重大死因に係る生活習慣病の共通経路であり、その進展要因とされている食生

活や運動習慣が与える影響を明らかにする必要がある。

本研究の目的は、人工知能による臨床エビデンスの統合と体系化を実現するため、各個人の生活情報や他のバイ

オマーカーなどの膨大な個人情報から、人工知能を用いて統合解析を進めることである。精度の高い生活習慣病合併症リスクエンジンの開発により、コホートにある食習慣をはじめとする食習慣や運動などについての問診などの個別化情報からの自動で診断し、科学的根拠に基づいた精密な予防や治療方針決定の実現化を目標とする。本研究の具体的な目的としては以下の2つが挙げられる。

1. 1 ヘルスケアの個別化アプローチ

“塩分摂取”を増量あるいは減量することで予後が改善する個人を同定する

人工知能を用いた従来の統計学的手法に比べ、精度の高い長期的死亡、高齢化に伴う疾患、QOL低下、メンタルヘルス、メタボリックシンドロームを予測するリスクエンジンの開発を行う。これは、これまでの古典的な統計学的手法によるものよりも高精度であることが求められる。さらにある種の生活習慣がなく、死亡リスクが高い特定の人々のうち、塩分摂取を促す(あるいは減らす)ことで、将来の死亡率を下げるができる集団を人工知能で特定する。この集団を特定することで、塩分摂取による健康効果をより一層促すことができる。したがって、本研究は費用対効果の高い、個別化された提案である。Precision nutrition は最適な健康のための栄養素の必要量、生活習慣が個人間で大幅に異なるという認識に起因する考え方である。このアプローチでは、遺伝学、エピジェネティクス、マイクロバイオーム、および他の栄養素、環境毒性物質、医薬品、疾患状態、および身体活動を含むがこれらに限定されない他の環境因子における差異が果たす役割を考慮する。つまり食事の影響が個人間に大きく異なるという異質性の度合いを検討するということである。また、個人の栄養と健康の間の最も長期的なリンクの1つは、マクロ栄養素の役割に主に焦点を当てた肥満と心臓代謝健康における食事の役割であること、かつ、微量栄養素と多量栄養素の両方がこれらの結果に影響し、神経学的、筋骨格系、生殖および免疫健康を含む他の健康結果に強い寄与因子であることを示している。これは一般的な疾患リスク発症モデルを作成しても人の行動変容にはつながらず費用対効果が低いという従来の疫学の問題点を解決するアプローチでもある。

1. 2 ヘルスケアの集団的アプローチ

“塩分摂取”に関するリアルタイムシステムティックレビュー・メタアナリシスの実現化

クリニカル・エビデンス(臨床エビデンス)の礎であるシステムティックレビュー・メタアナリシスを、人工知能を用い完全自動化、リアルタイム化を目指すものである。従来では人間が手作業で行なっている医学文献の評価や解析を、機械処理することにより論文数が膨大であっても高速に評価を行なうことができる。機械処理による高速化は現在と今後の潮流であり、本提案もその流れに沿うものである。すでに行なわれた予備実験では、十分な高速化が可能であるとの見込みが立っている。本研究成果は世界の全疾患領域の診療ガイドライン作成の効率化に大きく貢献する。さらに我々の住むアジアの日本語をはじめとする非英語言語も対応した人工知能プログラムを開発する。過去に発表された、塩分摂取に関する膨大な医学情報内に記載されたテキストの検索/閲覧の省力化をテキストマイニングの手法を用い可能とする。人工知能によるニューラルネットワークを介したディープラーニングには一般に膨大な教師データが必要となる、論文検索/閲覧支援プログラムの開発により、ディープラーニングのための教師データの作成の省力化の実現が期待される。過去の塩分摂取と生活習慣病に関連した膨大な臨床試験や疫学試験などの科学的根拠を統合解析し、臨床的社会的な位置づけを明確にして、効果的な予防に活用する。

2. 研究方法

2. 1 ヘルスケアの個別化アプローチ

“塩分摂取”量を変更することで予後が改善する個人を同定する

1. データ収集とデータ蓄積, 2. データ前処理・学習・推論エンジン生成, 3. 開発・実行・モニタリングが必要であり、まずは既存の参考となるコホートデータなどを用い、モデルの構築を目標とした。

研究対象: 日米大規模コホートデータを用いる。

米国民健康栄養調査(NHANES)は米国民の健康状態、栄養状態を定期的に調査し、面接による人口動態学的データ、社会経済学的状況、食事、健康状態の聞き取りのほか、医学的諸検査、身体測定を行っている。これらのデータは、主要疾患、疾患発症危険因子の有病率、栄養状態の把握とその健康増進、疾病予防との関連の検討

に供される。参加者の食品アンケート記録詳細、生活習慣病発症、死亡率に関連するデータを用いる。

統計解析: データはトレーニングデータ、検証データ、およびテストデータに分割する。トレーニング検証データを使用して、モデル構造のハイパーパラメータを調整する。テストデータはモデルテストと一般性テストに使用し Up sampling, SMOTE matching を用いる(図 1)。ディープラーニングによってアルゴリズムを作成する。検証データでパラメータを調整し、トレーニング/検証データでロジスティック回帰と深層学習を比較する。ディープラーニングまたはロジスティック回帰の最終決定バージョンアルゴリズムがテストデータに適用される。テストデータでアルゴリズムをテストし、再度データをテストする際のロジスティック回帰と深層学習の精度/ AUC の比較を行う。最終的にはオンライン上などでの実装を目標とし、誰もが自由に無料で利用できることが望まれる。

2.2 ヘルスケアの集団的アプローチ

“塩分摂取”に関するリアルタイムシステムティックレビュー・メタアナリシスの実現化

ニューラルネットワークを介した人工知能では、文章を読んだ人が重要だと判断する暗黙の基準を学習(ディープラーニング)し、多数の文書からその基準に沿ったものを抽出できる。教師データを学習する際に、教師データに含まれる単語ごとに教師データとの関連性と単語同士のつながりに関して伝達情報量を通して数値化し、ニューラルネットワークを介し、特徴量として利用している。教師データを学習した後、新たに評価用に投入されたデータにも同様の処理がなされ、教師データによって付与された特徴量により、それぞれの文章はスコア化され、そのスコア(=正解である予測確率)を基に教師データとの関連性を、塩分摂取量に関する臨床ガイドラインに採用されたシステムティックレビュー・メタアナリシス論文を評価した。(図 2)⁽¹⁾

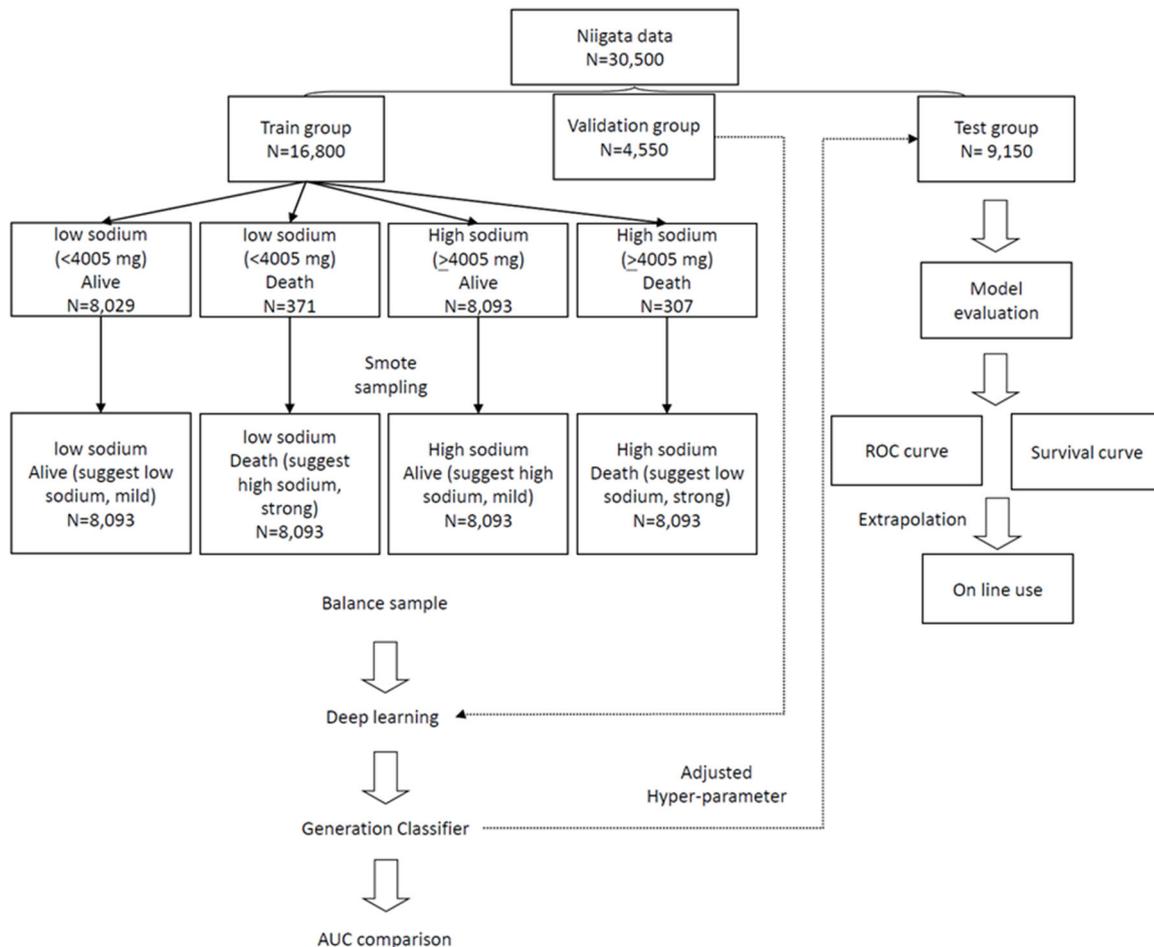
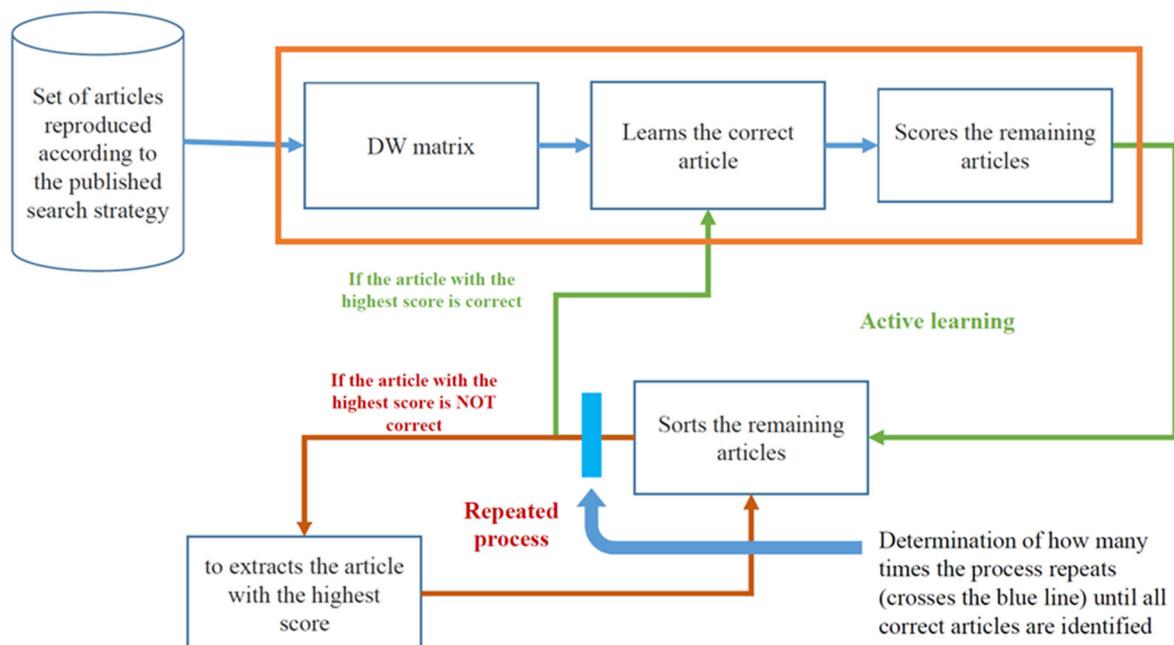


図 1. 解析モデル



Yamada et al. J Med Internet Res. 2020 Dec 30;22(12):e22422. doi: 10.2196/22422

図 2. 解析モデル

3. 研究結果

3. 1. ヘルスケアの個別化アプローチ

参加者の背景を表 1 に示す。参加者を Train group と Validation group, Test group に分割した。Train group を、塩分摂取量と、5 年追跡期間の中での死亡の有無の 4 グループに分割した。ADASYN Sampling し、仮想データ構築と Matching を行い、バランスをとった。Deep learning し、これに Validation group を用い、Hyper parameter を調整した。その後 Generation Classilfering し、Test group で Validation を行った。Logistic model と AUC について比較を行った。

テストグループデータの単純な比較 (= 集団レベルでの比較) では、塩分摂取量の大小による死亡率の差はなかった。1 日塩分摂取量平均以上を High sodium intake, 平均未満を Low sodium intake とした。(図 3)

そこで、構築したアルゴリズムを用いたシミュレーション解析の結果、塩分の摂取を増やす(減らす)ことで予後が改善すると推定された個人が塩分摂取量を変化させた場合に死亡リスクが低下する傾向を認めた。(図 4, 図 5)

3. 2. ヘルスケアの集団的アプローチ

“塩分摂取”に関するリアルタイムシステムティックレビュー・メタアナリシスの実現化

本研究では国内外の診療ガイドラインで実際に引用されたシステムティックレビュー・メタアナリシスを用いた検証を行い、人工知能アルゴリズムの実用化に向けた改良を行った。塩分摂取も含めた食事療法介入研究試験 (DPP) のメタアナリシス論文⁽²⁾で実際にシステムティックレビューに用いた論文を、正解論文(最終的に系統的レビューの対象となった論文)、非正解論文(系統的レビュー対象とならなかった論文)に分けた。人工知能が、教師データをもとに機械学習し、評価データにスコア付けした。結果 2 報の論文をランダムに抽出し、機械学習を行うことで、専門家の暗黙知を理解した。その後 AI が独自に付けた高スコア順に論文を閲覧した場合、全候補論文の 10% 未満の閲覧で、100% の正解論文をみれなく抽出することができ 90% の論文スクリーニングの省力化が得られた。(図 6)⁽¹⁾

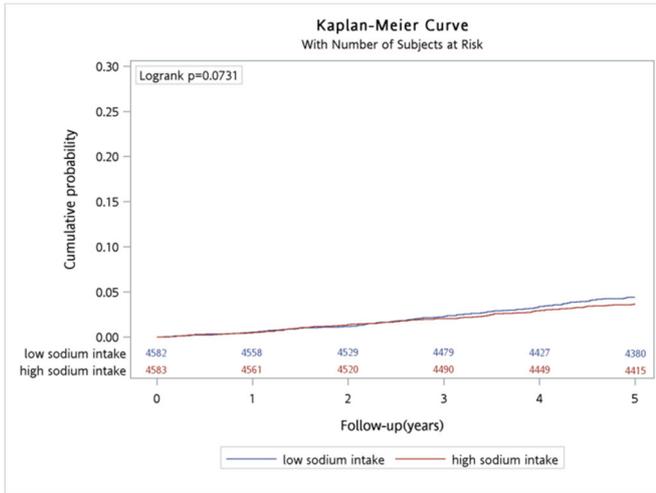


図 3. テストデータの塩分摂取の多少による 5 年死亡イベントの比較

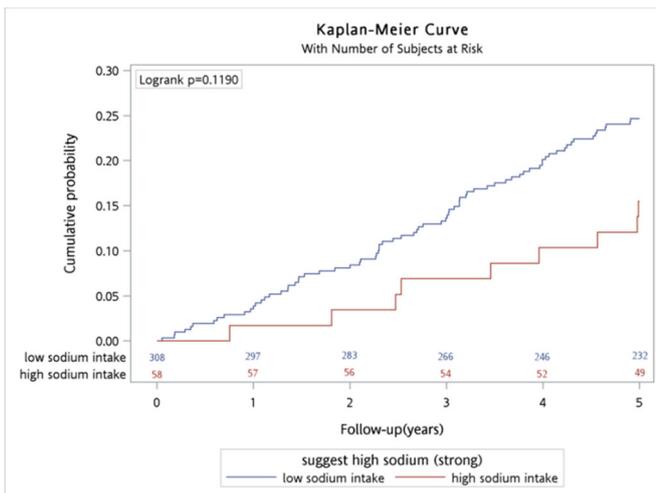


図 4. アルゴリズムを用いたシミュレーション解析(塩分摂取を増やすことが推奨される場合)

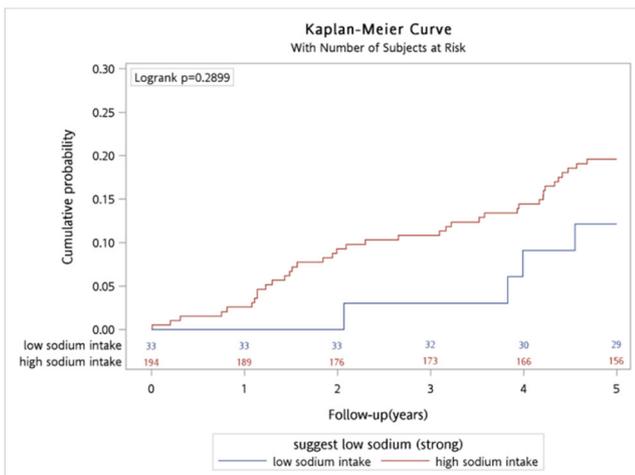


図 5. アルゴリズムを用いたシミュレーション解析(塩分摂取を減らすことが推奨される場合)

4. 考察と今後の課題

4. 1. ヘルスケアの個別化アプローチ

“塩分摂取”量を調整することで予後が改善する個人を特定する。

シミュレーションの結果、死亡率が下がるという傾向は捉えることができた。今後は塩分摂取量の設定や特徴量を変えていく。さらに、最適な特徴量を見出すためにデータについても増やしていく。かつ、死因別の解析をすすめる。Precision Nutrition の AI モデルを構築することで、リスクの高い人々に適切な塩分摂取の推奨を提案することができるようになる。これは個別化された費用対効果の高い提案である。

4. 2. ヘルスケアの集団的アプローチ

“塩分摂取”に関するリアルタイムシステムティックレビュー・メタアナリシスの実現化“

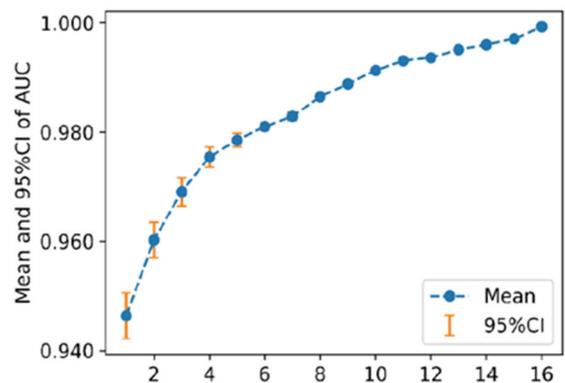
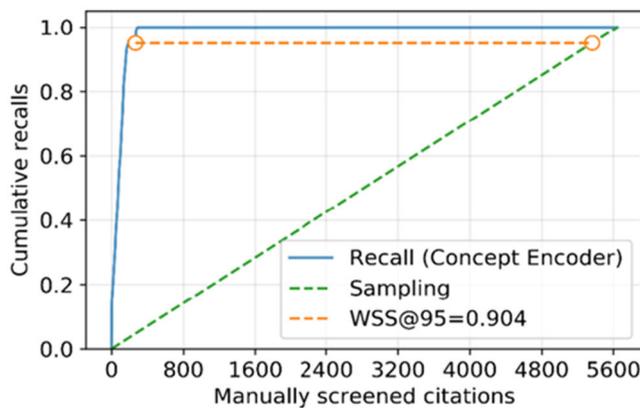
文献検索省力化に関しより検討を進める。この他我々は、論文の質的評価の高速化を目的としたソフト、PICORON-EBM (PICO and Risk Of bias Navigator for Evidence Based Medicine) の開発を行った。システムティックレビュー・メタアナリシスにおいて、多くの医学臨床論文の質(バイアスなど)に関して適切に評価することが求められる。しかし、個々の論文の質的評価には時間がかかることや時に見落としが生じることが問題であった。我々はこの質的評価時間を短縮し、見落としを防ぐためのテキストマイニングソフトウェア PICORON (PICO and Risk of Bias Navigator) を開発した。⁽³⁾PICORON に英文の臨床試験の論文を読み込ませることで、質的評価に必要な単語が自動でハイライトされ、論文の質的評価時間は 3 分の 1 まで短縮することができた(結果省略)。

また、医療情報からのデータ抽出ソフト T-Library の開発を行った。⁽⁴⁾医療情報(退院サマリ)の情報を項目ごとに自動で分割抽出し、エクセルデータ上の対応項目に分割され転記することができる。T-Library により退院サマリデータの転記にかかる時間は 5 分の 1 に省力化できた(結果省略)。このソフトにより数千数万のテキストベースの退院サマリなどのビッグデータを解析することが容易となり、自動更新される疾患リスクエンジンや人工知能を用いた解析に利用することが可能となる。

煩雑な作業を、専門家の暗黙知を学んだ人工知能が代行することで、人間はより創造的で生産的な仕事に従事することが可能となる。

ディープラーニングにより食事生活習慣介入研究試験 (DPP) のシステマティックレビューにおける、90% の論文スクリーニングの省力化が得られた。

Yamada et al. J Med Internet Res. 2020 Dec 30;22(12):e22422. doi: 10.2196/22422



PLOS MEDICINE

Diabetes Prevention Program Research Group
による検証

RESEARCH ARTICLE

Cardiometabolic Risk Factor Changes Observed in Diabetes Prevention Programs in US Settings: A Systematic Review and Meta-analysis

Lima Mudaliar^{1*}, Azadeh Zabetian³, Michael Goodman¹, Justin B. Echouffo-Tcheugui², Ann L. Albright², Edward W. Gregg³, Mohammed K. Ali^{1,3}



¹ Rollins School of Public Health, Emory University, Atlanta, Georgia, United States of America, ² Division of Endocrinology, Diabetes and Hypertension, Harvard Medical School, Boston, Massachusetts, United States of America, ³ Division of Diabetes Translation, National Center for Chronic Disease Prevention and Health Promotion, US Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America

* lmudaliar@gmail.com

図 6. 論文検索の省力化

6. 文献

- 1) Yamada T, Yoneoka D, Hiraike Y, et al, Deep Neural Network for Reducing the Screening Workload in Systematic Reviews for Clinical Guidelines: Algorithm Validation Study. J Med Internet Res 2020 Dec 30;22(12):e22422.
- 2) Mudaliar U, Zabetian A, Goodman M, Echouffo-Tcheugui JB, Albright AL, Gregg EW, et al, Cardiometabolic risk factor changes observed in diabetes prevention programs in US settings: a systematic review and meta-analysis. PLoS Med 2016;Jul;13(7):e1002095
- 3) Yamada T, Kondo Y, Momosaki R, PICO and Risk Of bias Navigator for Evidence Based Medicine. PeerJ Preprints 2019;7:e27684v1 DOI 10.7287/peerj.preprints.27684v1
- 4) Yamada T, Kondo Y, Momosaki R, Automated data extraction software for medical summary using text mining (T-Library) PeerJ Preprints 2019;7:e27685v1 DOI 10.7287/peerj.preprints.2768

Integration and Systematization of Clinical Evidence for Salt Intake by Utilizing Artificial Intelligence

Tomohide Yamada^{1,2)}, Nobuhiro Shojima²⁾, Chia-lin Lee³⁾, Hisashi Noma⁴⁾, Toshimasa Yamauchi²⁾

¹⁾King's College London, London, UK, ²⁾The University of Tokyo, Tokyo, Japan, ³⁾Taichung Veterans Hospital, Taichung, Taiwan, ⁴⁾Institute of Mathematical Statistics, Tokyo, Japan

Summary

【Background】Evidence-based medicine (EBM) involves determining treatment that matches the needs of each patient by integrating the best and latest available "scientific evidence" and "clinical skills". Systematic review and meta-analysis refers to the process of searching databases and performing statistical analysis to integrate the results of multiple independent studies conducted in the past. These results obtained provide the highest quality evidence, which has become the foundation of various clinical guidelines. The objectives of this project are to reduce the time required to perform a systematic review by employing artificial intelligence (AI) and to improve the precision of the method.

We aimed to identify particular people those who have benefited from salt intake at high risk of 5-year mortality by Artificial Intelligence. This approach takes into account the role played by differences in genetics, epigenetics, the microbiome and other environmental factors including but not limited to other nutrients, environmental toxicants, pharmaceuticals, disease status, and physical activity.

【Methods】The objectives and planned outcomes of this project are to reduce the time required to perform a systematic review by employing artificial intelligence (AI) and to improve the precision of the method.

【Results】By using AI software, the aim is to make comprehensive searches of medical articles for reviews 10 times faster compared to the current speed.

【Conclusion】The results of this study could provide significant benefit by facilitating the acquisition of new evidence for clinical guidelines in all disease fields as well as salt intake.

Key words: Artificial intelligence, Systematic Review, Life-related diseases.