

国立大学法人東京大学

東京大学と Google との AI を活用した 多遺伝子リスクスコアモデル研究について協業

-日本における疾患管理や予防医療発展への貢献を目指す-

国立大学法人東京大学と Google 日本法人は AI を活用した多遺伝子リスクスコア（PRS¹: Polygenic Risk Score）モデル研究の協業に関して契約を締結しました。本協業は、多様な社会問題の解決と AI が人と相利共生して生み出す未来社会「AI相利共生未来社会」の実現に向けて、東京大学と Google が2020年10月に締結したパートナーシップの一環です。

Google は、これまでに、ゲノム配列データにおける変異やバリエーション²を特定するプロセスであるバリエーションコーリングの精度を向上させる AI モデルを開発し、2018 年に [Nature Biotechnology](#) で発表しています。また、ゲノム配列データから目の色や特定の病気などの身体的特徴をより正確に予測できるモデルや、これを利用して特定の特徴と関連するバリエーションを発見するモデルも開発しています。

東京大学医科学研究所では、これまで[バイオバンク・ジャパン](#)に蓄積したデータの解析を通じて肥満症、血液検査、2型糖尿病、心筋梗塞、心房細動等の数々のゲノムワイド研究において非欧米系集団で最大の貢献をしてきました。

今回の協業では、Googleが開発した ヨーロッパ地域の人々を対象としたPRS を予測するAIモデルを、日本人を対象に検証、改良し、東京大学のゲノムワイド研究の知見をもって評価することで、日本における疾患管理や予防医療の発展への貢献を目指します。

現在、多くの遺伝学的研究は、ヨーロッパ地域に祖先をもつ個人を対象にしています。それらの人々は世界人口の 4 分の 1 以下であるにもかかわらず、既存の GWAS（ゲノムワイド関連解析）研究³の 4 分の 3、そして PRS 研究の 3 分の 2 以上は彼らが対象です。病気やその治療法に関する理解は、しばしばこれらの研究に基づいており、ヨーロッパ地域に祖先をもつ人々を対象に開発された PRS の予測性能は他の集団では低くなること

¹ PRS は、多くのバリエーションの影響を1つのスコアにまとめ、バリエーションに起因する疾患リスクを推定するために使用可能。個人がある特定の病気にどの程度かかりやすいかを示すもので、本人や医療従事者が、病気の予防や、早期発見、治療の計画に役立つ可能性があると考えられる。

² 生物の設計図であるゲノムには個人差があり、例えばヒトでは個人間で99.5%以上は同一であるが、残りの部分では違いがある。個人ごとに異なるような生殖細胞系列ゲノム配列上の位置をバリエーションと言う。

³ 全ゲノムのバリエーションを解析し、疾患などのヒトの特徴と関連するものを網羅的に検出する手法。世界初の GWAS は 2001 年に日本の理化学研究所から報告されている。一般にPRS はGWAS結果を利用して構築する。

があるなど、集団間での精度の違いが影響を与える可能性もあります。本研究は、バイオバンク・ジャパンの包括的なデータと、Google の遺伝統計学とディープニューラルネットワーク⁴の専門性を合わせることで、多くの集団に対する遺伝学的研究の有用性を広げる機会を提供します。

今回の協業では、東京大学医科学研究所・所長オフィスのバイオバンク・ジャパンのデータを使用し、機械学習モデルのパフォーマンス分析を行います。これらのデータは、本協業での利用のみを目的として、高水準のセキュリティを有する Google Cloud の日本国内に設置したサーバーに暗号化した上で保管されます。

東京大学大学院新領域創成科学研究科 教授の鎌谷洋一郎氏は、本協業に関して以下のよう
に述べています。

「バイオバンク・ジャパンは非欧米系として世界最大級の疾患バイオバンクであり、PRS研究の地域間の予測性能差を[最初に大規模に証明した研究](#)でも主要な役割を果たしました。今回のデータ利活用と Google の先進的な AI 技術によって、平等で公正な予測性能を持ち、全世界で健康と福祉の増進につながる PRS の開発につながることを期待しています。また、バイオバンク・ジャパンは今後も医療の革新を目指して産業界を含めた更なるデータ利活用の推進をおこなっていきます。」

本日発表した協業が、日本、そして世界の疾病管理や予防医療の発展に貢献するよう願っています。

【本件に関する報道関係者からの問い合わせ先】

- ・東大-Google間のパートナーシップについて

国立大学法人東京大学 産学協創部

e-mail : kyoso-info.adm@gs.mail.u-tokyo.ac.jp

※メールの件名の冒頭に【東大Google 多遺伝子リスクスコア】と記載してください

- ・本協業で実施される研究について

新領域創成科学研究科 教授 鎌谷洋一郎

e-mail : kamatani.yoichiro@edu.k-u-tokyo.ac.jp

- ・バイオバンク・ジャパン (BBJ) について

BBJ事務局 東京大学医科学研究所内

e-mail : bbjkoho@biobankjp.net

⁴ 一般に統計解析では事前に目的変数と説明変数との関係性を数理モデルとして記述する必要があるのに対し、機械学習ではデータとその学習ルールだけを与えれば、変数間の複雑な関係性・規則性をヒトが指定しなくても機械的に学習される。ディープラーニングは機械学習方法の一種で、深層学習とも呼ばれる。ニューラルネットワークと呼ばれる生物の神経回路を模倣した機械学習モデルを何層も繰り返すことで、さまざまな特徴量を作り出しつつ、学習がうまくいくようなさまざまな工夫を重ねることで、高精度な予測を実現する。