PRESS RELEASE

報道解禁（日本時間）：4 月 20 日 15 時（21 日朝刊）

配信先：大学記者会（東京大学） 文部科学記者会 科学記者会 岩手県庁教育記者クラブ

2026 年 4 月 16 日

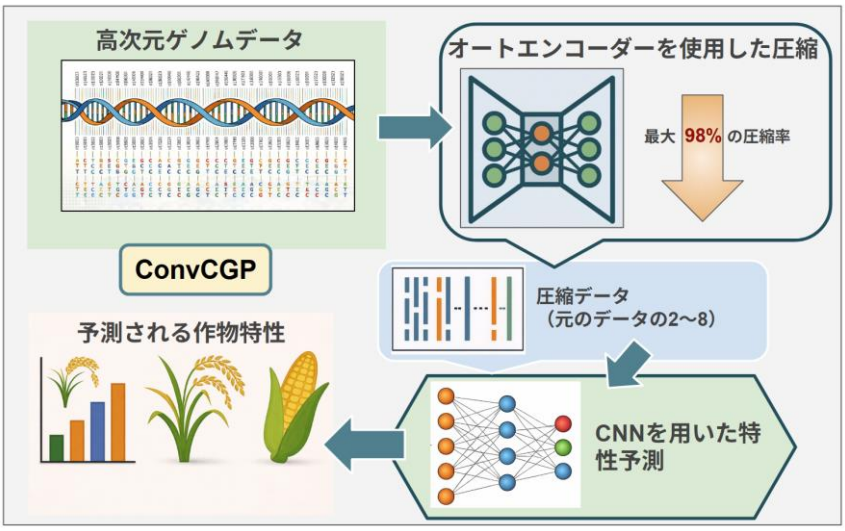
東京大学
岩手大学

DNA 情報を 98%圧縮し、作物の性質を高精度予測

―AI により計算時間を大幅短縮、品種改良を加速―

発表のポイント

- 数百万～1,000 万以上の遺伝マーカーを含むゲノムデータを、**最大 98%圧縮しても高い予測精度を維持する手法**を開発しました。
- データ圧縮により、**計算時間とデータ保存コストを大幅に削減**しました。
- イネおよびトウモロコシのデータで、従来手法よりも**高い予測精度**を達成しました。
- DNA 情報から有望な品種を効率的に選びだす技術として、**品種改良の加速に貢献**します。



本研究の概要

膨大な DNA 情報を AI で圧縮し、その圧縮データから作物の性質（収量や草丈など）を予測することで、有望な品種の選抜を効率化する。

概要

東京大学大学院農学生命科学研究科の雷帆タンジラ特任助教と岩田洋佳教授らの研究グループは、作物のゲノムデータを大幅に圧縮しながら、収量や草丈などの性質を高精度に予測できる新しい深層学習手法「ConvCGP」を開発しました。

近年の品種改良では、**数百万から 1,000 万以上の遺伝マーカーを含むゲノムデータ**を用いて、作物の性質を予測する「ゲノミック予測」が活用されています。しかし、データの大規模化により、計算時間やデータ保存の負担が大きな課題となっていました。

本研究では、重要な遺伝情報を保ったままデータを圧縮し、その圧縮データから作物の性質を予測する手法を開発しました。その結果、元のデータの**わずか 2%程度に圧縮しても高い予測精度を維持できる**ことを示しました。

発表内容

背景

品種改良では、収量や品質、環境ストレスへの耐性などに優れた品種を効率的に選ぶことが重要です。近年は、DNA 情報（ゲノムデータ）をもとに作物の性質を予測する「ゲノミック予測[1)]」が広く用いられており、実際に栽培する前の段階で有望な個体を選抜できるようになっています。これにより、品種改良の期間短縮と効率化が進んでいます。

一方で、ゲノミック予測では、数百万から 1,000 万以上の遺伝マーカーを含むゲノムデータ[2)]を扱うため、計算時間やデータ管理の負担が大きく、特に大規模な育種プログラムにおいて課題となっていました。

研究成果の内容

本研究では、この課題を解決するため、深層学習[3)]を用いた新しい手法「ConvCGP」を開発しました。本手法は、まずオートエンコーダによってゲノムデータを圧縮し、その後、畳み込みニューラルネットワーク（CNN）によって作物の性質を予測する二段階の構造を持ちます。これにより、重要な遺伝情報を保持したまま、効率的な予測が可能となります。

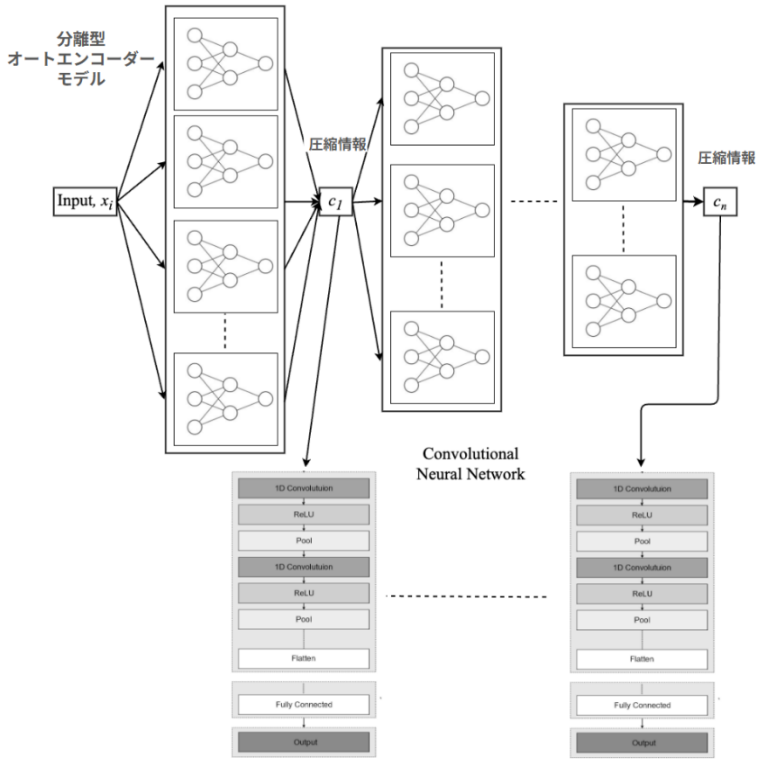


図 1：提案手法「ConvCGP」の概要

ゲノムデータを AI で圧縮（オートエンコーダ）し、その圧縮データから AI（CNN）によって作物の性質を予測する二段階の手法。

イネおよびトウモロコシの大規模データを用いた検証では、**データを 93～98%圧縮（元データの 2～7%）しても、高い予測精度を維持できる**ことが確認されました。特に、約 70 万マーカーのイネデータや約 1,170 万マーカーのトウモロコシデータにおいても、圧縮後のデータから高精度な予測が可能でした。

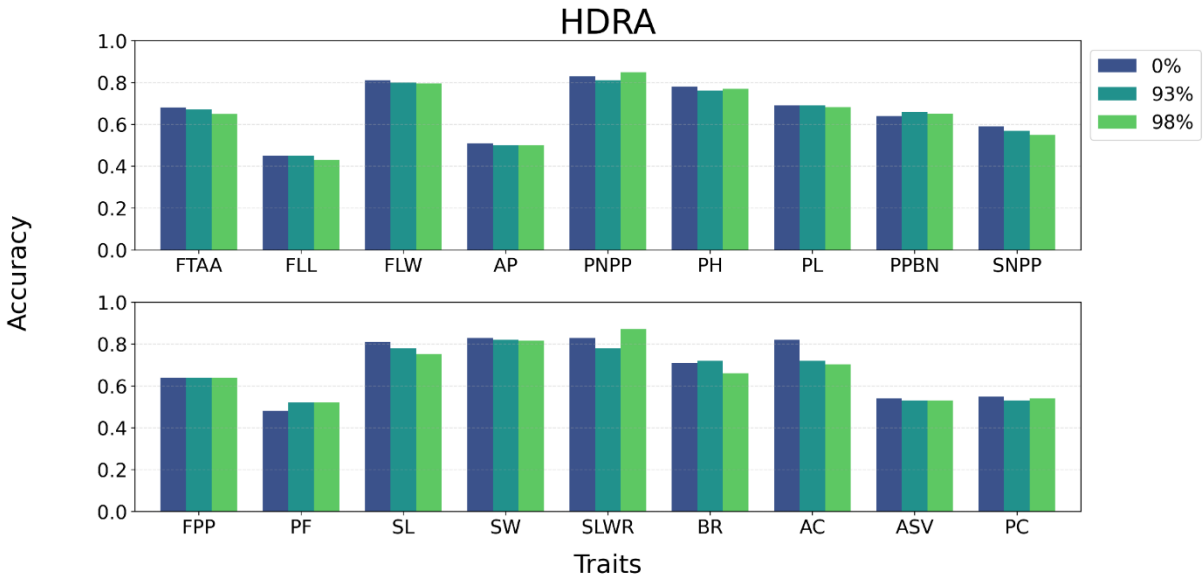


図 2:ゲノムデータ圧縮後の予測精度(イネ)

ゲノムデータを最大 98%圧縮しても、収量や草丈などの性質の予測精度がほぼ維持されることを示す。

FTAA:アーカンソーにおける出穂期, FLL:止葉長, FLW:止葉幅, AP:芒の有無, PNPP:株当たり穂数, PH:稈長, PL:穂長, PPBN:一次枝梗数, SNPP:穂当たり粒数, FPP:穂当たり小花数, PF:稔実率, SL:粒長, SW:粒幅, SLWR:粒長幅比, BR:いもち病抵抗性, AC:アミロース含量, ASV:アルカリ崩壊値, PC:タンパク質含量

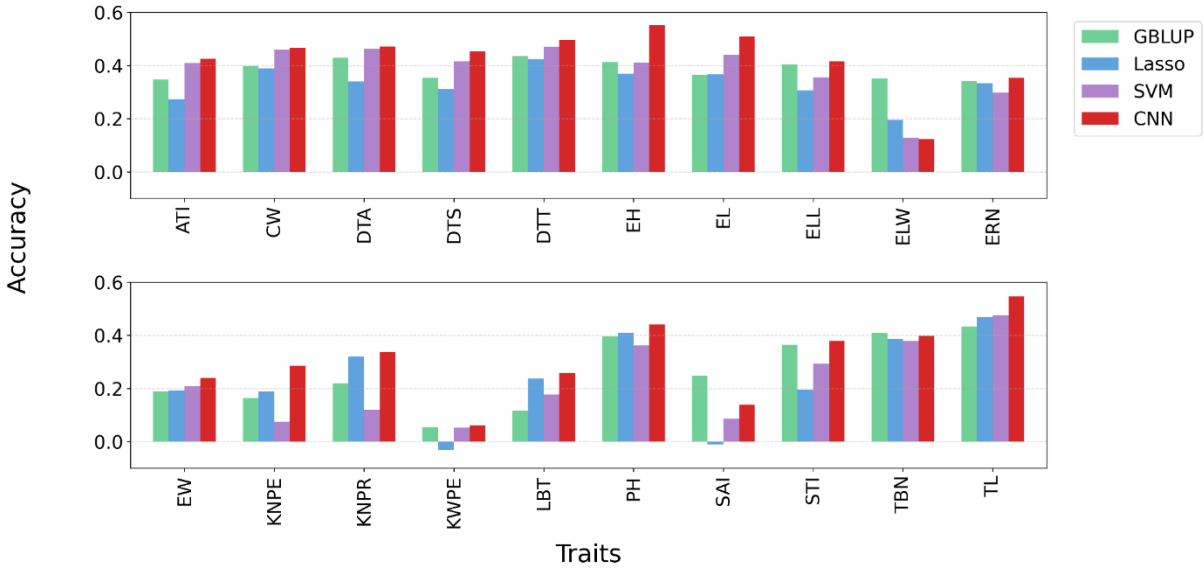


図 3:従来手法との予測精度の比較(トウモロコシ)

ConvCGP は、GBLUP や Lasso などの従来手法と比べ、多くの性質において高い予測精度を示す。

ATI:葯-雄穂間隔, CW:穂芯重, DTA:雄穂開花日数, DTS:絹糸抽出日数, DTT:雄穂出穂日数, EH:着雌穂高, EL:穂長, ELL:穂位葉長, ELW:穂位葉幅, ERN:穂列数, EW:穂重, KNPE:穂当たり粒数, KNPR:列当たり粒数, KWPE:穂当たり粒重, LBT:不稔先端長, PH:草丈, SAI:絹糸-葯間隔, STI:絹糸-雄穂間隔, TBN:雄穂分枝数, TL:雄穂長

また、計算効率も大幅に向上しました。例えば、イネのデータでは予測に要する時間が約 2 分 50 秒から約 14 秒へ短縮され、大規模データでは 1 日以上かかっていた計算が数十分程度まで短縮されました。

さらに、GBLUP や Lasso、サポートベクターマシンなどの従来手法と比較して、多くの性質においてより高い予測精度を示しました。開花時期や草丈、病害抵抗性など、複雑な遺伝的背景を持つ性質に対しても安定した予測が可能であることが確認されました。

本手法は圧縮率を用途に応じて調整できるため、大規模スクリーニングから精密な品種設計まで柔軟に対応できます。

社会的意義・今後の展開

本研究で開発された技術により、DNA 情報から作物の性質を迅速に予測し、有望な品種を効率的に選抜することが可能になります。これにより、品種改良の高速化とコスト削減が期待されます。また、本技術はデータ処理に伴う計算資源の削減にもつながるため、データ駆動型農業の効率化や持続可能な農業（GX：グリーントランスフォーメーション）の推進にも貢献すると期待されます。今後は、環境データとの統合や他作物への応用を進めることで、より高度な予測と実用化を目指します。

発表者・研究者等情報

東京大学
大学院農学生命科学研究科
雷帆　タンジラ　特任助教（Tanzila Islam）
岩田　洋佳　教授（Hiroyoshi Iwata）

岩手大学
理工学部
金　天海　准教授（研究当時）（Chyon Hae Kim）
（現 スカイオーシャンテクノロジー株式会社 取締役 CTO）
農学部・次世代アグリイノベーション研究センター
下野　裕之 教授（Hiroyuki Shimono）
理工学部・次世代アグリイノベーション研究センター
木村　彰男 教授（Akio Kimura）

論文情報

雑誌名：The Plant Genome
論文タイトル：ConvCGP: A Convolutional Neural Network to Predict Genetic Values of Agronomic Traits from Compressed Genome-wide Polymorphisms
著者：Tanzila Raihan, Chyon Hae Kim, Hiroyuki Shimono, Akio Kimura, Hiroyoshi Iwata
DOI：https://doi.org/10.1002.TPG2.70223

用語解説

1）ゲノミック予測：DNA 情報（ゲノムデータ）をもとに、作物の性質（収量や草丈など）を予測する手法。

2) ゲノムデータ：DNA 上の多数の遺伝マーカーの情報。作物では数百万から 1,000 万以上のマーカーが利用される。
3) 深層学習：データの特徴を自動的に学習する人工知能の一種。本研究では、データ圧縮と性質の予測に用いられる。

注意事項（解禁情報）

日本時間 4 月 20 日 15 時 （UTC 4 月 20 日午前 6 時）以前の公表は禁じられています。

研究助成

本研究は、JSPS 科研費（JP19H00938、JP22H02306）および JST AI-ENGENE プログラム（JPMJMS25E3）の支援を受けて実施されました。

問合せ先

（研究内容については発表者にお問合せください）
東京大学 大学院農学生命科学研究科
教授 岩田 洋佳 （いわた ひろよし）
E-mail: hiroiwata[at]g.ecc.u-tokyo.ac.jp

東京大学 大学院農学生命科学研究科・農学部
事務部 総務課総務チーム広報情報担当
Tel:03-5841-5484 E-mail:koho.a[at]gs.mail.u-tokyo.ac.jp

岩手大学 法人運営部 総務広報課広報グループ
Tel:019-621-6015 E-mail:kkoho[at]iwate-u.ac.jp

※[at]は@に変えて入力ください。