

# アンサンブル学習を用いた造血幹細胞移植予後予測モデルの開発

## —機械学習を用いた新規生存時間解析手法の実装—

### 概要

京都大学大学院医学研究科 高折晃史教授、諫田淳也同助教、岩崎惇同研修員、ネクスジェン株式会社 中島正和氏、宮西正憲氏、宮塚功氏、Le My An 氏らの共同研究グループは、新規の造血幹細胞移植予後予測モデルを開発しました。

同種造血幹細胞移植においては、患者年齢や疾患等の患者因子、並びに HLA 一致度やドナーソース等のドナー側因子を含む数多くの因子が予後に関わる事が知られています。本研究グループは、従来、予後予測に用いられてきた統計学的手法、並びに機械学習手法で得られた予測モデルを、アンサンブル学習（注1）を用いて融合し、新たな予測モデルを開発しました。京都大学血液内科関連病院の移植データを用いた解析では、同モデルは従来の手法と比較し、全生存、再発、GVHD（注2）、並びに上記を考慮した重篤な病的状態を伴わない生存の指標である GVHD-free, Relapse-free Survival (GRFS)（注3）、いずれにおいても優れた予測精度 (C-index、注4)を示しました。各症例に適したドナーの選択や、治療に対する患者の理解に貢献する事が期待されます。

本成果は、2021年12月21日に米国の国際学術誌「Blood Advances」にオンライン掲載されました。



## 1. 背景

造血幹細胞移植領域での予後予測には、Cox 回帰分析などの統計学的手法が用いられてきました。近年、機械学習を基にした手法が開発されてきており、ゲノム情報等新しい因子も含めて、多数の因子が関与する造血幹細胞移植の予後予測に有用性が示されていますが、従来の方法と比較して予測精度を高める手法の開発は十分に進んでいません。そこで本研究グループでは、従来法より精度の高い予後予測アルゴリズムを開発する事を目的として本プロジェクトを立ち上げました。

## 2. 研究手法・成果

従来、予後予測に用いられてきた統計学的手法、並びに既存の機械学習手法で得られた予測モデルを、アンサンブル学習を用いて融合し、新たな予測モデルを開発しました。同モデルは従来手法と比較し、全生存、再発、GVHD といった個別の予後に加えて、上記を考慮した重篤な病的状態を伴わない生存の指標である GVHD-free, Relapse-free Survival (GRFS)、いずれの予後においても優れた予測精度(C-index)を示しました (図1)。本モデルを使用する事により、個々の患者に合わせた移植方法の選択や、移植の導入前の説明の段階でより良い患者理解を得る事に貢献すると考えています。

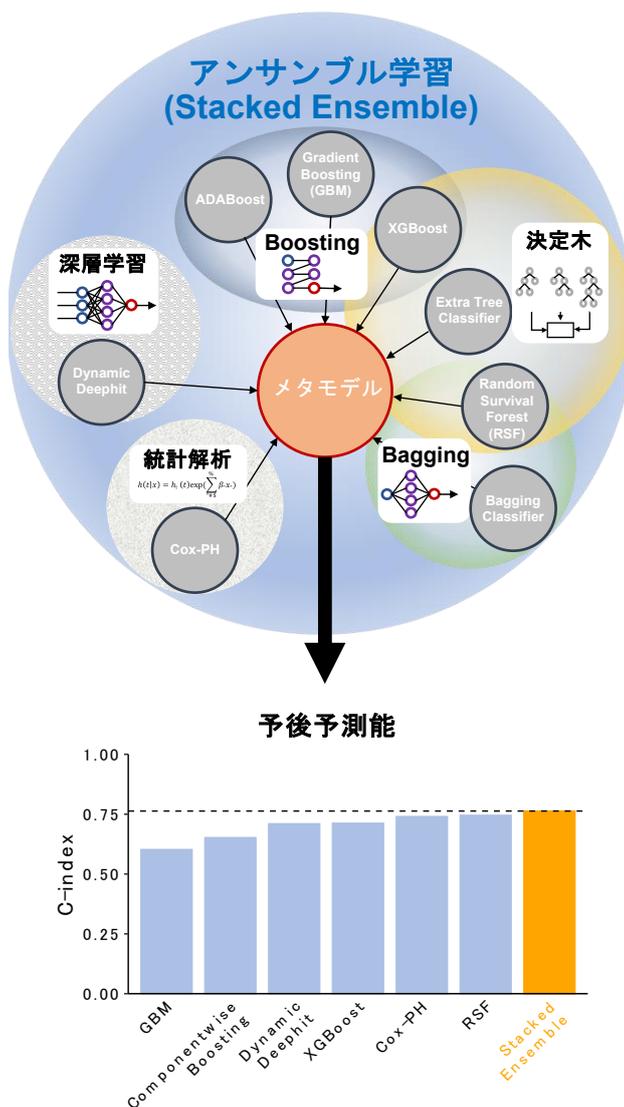


図1 アンサンブル学習および他モデルにおける予測精度

### 3. 波及効果、今後の予定

多数の因子を基に正確な予後予測ができる事は、個々の患者に適した移植方法を提供する個別化医療の実現に貢献する事が期待されます。ウェブツールの開発も並行して進めており、本研究で得られたモデルから移植後の予後を臨床の現場で予測し活用していただける準備も整えています。

今後の方向性としては、本研究では、京都大学関連病院の症例に限って解析を実施しているが、日本国内でも施設によって移植方法に違いはあり、また、近年新たな治療法の開発・普及に伴い、ドナー選択を始めとした移植方法も変化してきている事から、今後もより実臨床に最適化したモデルにしていく事が必要だと考えます。全国のデータ、あるいは海外のデータも含めた解析、更には実際に臨床の現場で応用した内容を前向きに本研究で得られたアルゴリズムに学習させる事で、より洗練された予測モデルの構築に繋がると考えています。

### 4. 研究プロジェクトについて

本研究は、国立研究開発法人日本医療研究開発機構（AMED）の医療研究開発革新基盤創成事業（CiCLE）の支援を受け、課題名「造血幹細胞の体外増幅技術の開発と移植医療への応用」として実施されたものです。

#### <用語解説>

（注1）アンサンブル学習：より優れた予測精度を示すモデルを示すために、個々の学習器に基づき作成されたモデルを融合して新たなアルゴリズムを作成する手法。本研究においては、複数の統計学的手法・機械学習アルゴリズムを、スタッキングと呼ばれるアンサンブル学習方法を用いて融合し、新たなモデルを作成している。

（注2）GVHD：移植片対宿主病。移植したドナーの免疫細胞がレシピエント（患者）の細胞を標的として働き、炎症を引き起こす事で起こる移植後に特徴的な免疫学的合併症。

（注3）GVHD-free, Relapse-free Survival：死亡、再発、重症急性GVHD（グレード3以上）、全身型慢性GVHDのいずれかの発症をエンドポイントとする、重篤な病的状態を伴わない移植後の生存を測定するための複合エンドポイント。

（注4）C-index：ある予後の発祥に関して、予測モデルが正しく判別する能力を測る指標として用いられる。C-indexは0から1までの値を取り、C-index = 1であれば完全に予測が一致する事を示し、C-index = 0.5の場合には、モデルが予後の予測に適合していないと判断できる。

#### <研究者のコメント>

アンサンブル学習を造血幹細胞移植の予後予測に用いて、より優れた予測を実現できる可能性を示した事で、機械学習の医学分野への直接的な応用に繋がる知見となる事を期待しています。今後、国内・海外の既存のデータや前向きに収集したデータを用いてモデルを発展させる事で、より実臨床に即したモデル構築を進められ、個々の患者に最適の治療を提供し、患者医師間で治療への理解を共有する良いツールとなると考えています。

#### <論文タイトルと著者>

タイトル：Establishment of a Predictive Model for GvHD-free, Relapse-free Survival after Allogeneic HSCT using Ensemble Learning

(アンサンブル学習を用いた同種造血幹細胞移植後 GRFS 予測モデルの確立)

著者：岩崎惇、諫田淳也、新井康之、近藤忠一、石川隆之、上田恭典、今田和典、赤坂尚司、米澤昭仁、野吾和宏、直川匡晴、安齋尚之、森口寿徳、北野俊行、伊藤満、有馬靖佳、竹岡友晴、渡邊光正、平田大二、浅越康助、宮西正憲、宮塚功、Le My An、高折晃史

掲載誌：Blood Advances DOI：10.1182/bloodadvances.2021005800