

P546 MACHINE LEARNING AND RESPONSE TO INDUCTION CHEMOTHERAPY IN PATIENTS WITH ACUTE MYELOID LEUKEMIA

Topic: 4. Acute myeloid leukemia - Clinical

Heba Elhaddad^{*1,2}, Claudia Chiriches^{1,3}, Katie Watts⁴, Peter Giles⁵, Amanda Gilkes^{1,3}, Martin Ruthardt^{1,3}

¹Division Of Cancer And Genetics, Hematology Department, School Of Medicine, Cardiff University, Cardiff, United Kingdom; ²Clinical Pathology Department, Mansoura University, Mansoura, Egypt; ³Experimental Cancer Medical Centre (Ecmc) Cardiff, School Of Medicine, Cardiff University, Cardiff, United Kingdom; ⁴Division Of Cancer And Genetics, School Of Medicine, Cardiff University, Cardiff, United Kingdom; ⁵Wales Gene Park, Cardiff University, Cardiff, United Kingdom

Background:

The toxic effects of induction chemotherapy for acute myeloid leukemia (AML) are only worth it if a patient responds. Until now, no validated scoring system can accurately predict patients' responses to chemotherapy. Most prediction approaches are complicated as they utilize various cytogenetic and molecular abnormalities. This is not only imprecise but also time-consuming. Walter et al. used a scoring system with an area under the curve (AUC) of 0.64 to 0.69 to predict resistance to chemotherapy. Ng et al. incorporated mutational data with other covariates to get concordance statistics (C-statistics) of 0.65 to 0.8 in response prediction, which was biased because death and relapse affected the C-statistics.

Machine learning (ML) aims to make accurate predictions based on existing data. Random Forest (RF) is an ML algorithm widely used to make predictions related to cancer, like progression and survival. In AML, RF has been recently applied in predicting patients' overall survival (OS). It achieved a prediction efficiency of AUC 0.59 to 0.64, better than the ELN risk stratification. Therefore, we focused on predicting the response to induction therapy based on mutational data, as induction failure can worsen patients' prognosis due to treatment delay and exposure to unnecessary toxicity.

Aims:

We applied ML to investigate whether targeted sequencing contained the relevant information to predict the patients' response to induction chemotherapy.

Methods:

In this study, sequencing data of 111 genes of a cohort of 1552 AML patients from the AML NCRI study group were retrospectively analyzed using RF. Following intense induction chemotherapy, patients were stratified into responders achieving complete remission (CR) or resistant not achieving CR. The CR rate in our dataset was around 80%, matching rates reported in the literature. However, 47% of patients in the CR group relapsed during follow-up.

Sequencing was performed on an Illumina platform. All analyses were done using R. The cohort was randomly split into training (70%) and test (30%) sets. Our starting dataset needed to be more balanced, having 80% CR vs. 20% resistance. Therefore, the training set was balanced using SMOTE before RF analysis.

Results:

As we aimed to distinguish responders from resistant patients, we first tested different relapse time thresholds to decide whether to include the relapsed patients in the CR or the resistant group. RF revealed that the best

Copyright Information: (Online) ISSN: 2572-9241

© 2023 the Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the European Hematology Association. This is an open access Abstract Book distributed under the Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) which allows third parties to download the articles and share them with others as long as they credit the author and the Abstract Book, but they cannot change the content in any way or use them commercially.

Abstract Book Citations: Authors, Title, HemaSphere, 2023;7(S3):pages. The individual abstract DOIs can be found at <https://journals.lww.com/hemasphere/pages/default.aspx>.

Disclaimer: Articles published in the journal HemaSphere exclusively reflect the opinions of the authors. The authors are responsible for all content in their abstracts including accuracy of the facts, statements, citing resources, etc.

prediction accuracy could be obtained when all relapsed patients were included in the CR group. Patients who died within 28 days of starting therapy were excluded, as early death introduced bias to RF analysis.

RF model with balanced data achieved 93% accuracy in predicting the training set (AUC 0.9). When the RF model was used to predict the treatment outcome of the test set (a new subset of data), it attained 79% accuracy, AUC 0.676, 87% sensitivity, 44% specificity, 86% positive predictive value (PPV), and 46% negative predictive.

The effect of data balancing is reflected on the prediction of the test set (after balancing: accuracy increased from 73% to 79%, and specificity increased from 19% to 44%), resulting in higher prediction accuracy for the minority (resistant) group.

Summary/Conclusion:

Targeted sequencing data obtained during diagnosis include the information for RF to predict the patients' response to induction chemotherapy. RF is a practical, rapidly developing algorithm that can be incorporated into future technologies to obtain high response prediction accuracy in AML patients

Copyright Information: (Online) ISSN: 2572-9241

© 2023 the Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the European Hematology Association. This is an open access Abstract Book distributed under the Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) which allows third parties to download the articles and share them with others as long as they credit the author and the Abstract Book, but they cannot change the content in any way or use them commercially.

Abstract Book Citations: Authors, Title, HemaSphere, 2023;7(S3):pages. The individual abstract DOIs can be found at <https://journals.lww.com/hemasphere/pages/default.aspx>.

Disclaimer: Articles published in the journal HemaSphere exclusively reflect the opinions of the authors. The authors are responsible for all content in their abstracts including accuracy of the facts, statements, citing resources, etc.