A Machine-Learning Model to Facilitate Individual-level Risk Screening in Patients with Myelodysplastic Syndromes in Routine Clinical Practice

RWD52

Scan for more information

VP Vaidya,[1] V Priya,[1] D Parmar,[1] R Yan,[1] RK Das,[1] M Haririfar,[1] P McMahon,[2] M Williamson,[2] C Hogea[2]

[1]ConcertAI, Cambridge, MA, USA.
[2]Gilead Sciences, Inc., Foster City, CA, USA.

Key Findings

- Automated, non-invasive machine learning (ML)–based methods can be employed towards regular risk-screening for patients with MDS utilizing information commonly collected in routine practice.
- Such tools could help prompt timely use of existing clinical scoring systems such as the International Prognostic Scoring System (Revised, IPSS-R; Molecular, IPSS-M), which involve more complex invasive testing.
- Model performance was comparable to established risk classification tools like IPSS-R and IPSS-M for overall survival (OS) prediction **(Table)**.

Conclusions

A data-driven ML model was developed that can reasonably predict the OS for patients with MDS based on individual patient characteristics.

This was achieved through integration of comprehensive EHR structured patient-level data coupled with temporal feature engineering to capture the patient clinical trajectory **(Figure 1)**.

Models like the one shown have the potential to alert for risk and cue confirmatory testing for timely patient care management. Further validation and evaluation in clinical practice are needed.

Introduction

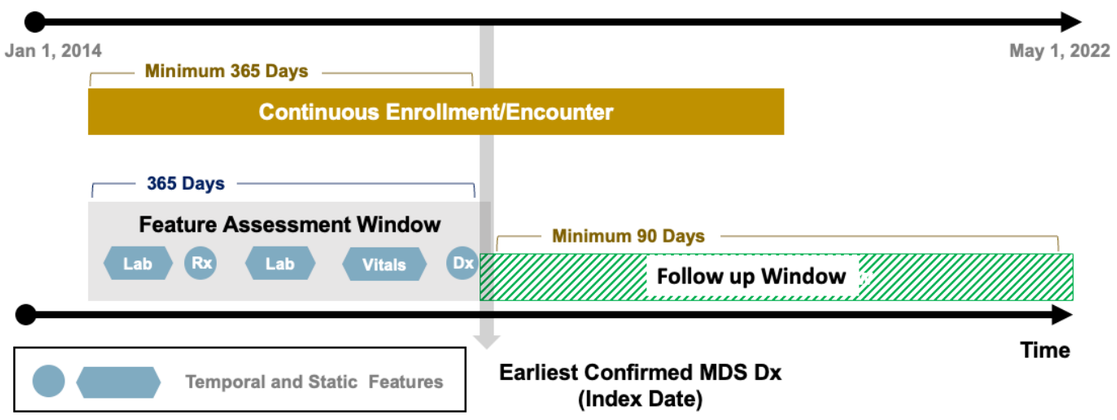
- Myelodysplastic syndromes (MDS) represent a heterogeneous group of myeloid neoplasms with risk of progression to acute myeloid leukemia (AML).
- Patient prognosis, management and outcomes differ based on risk category; higher-risk MDS patients have significantly lower OS and faster progression to AML.
- Early presentation and symptoms can be non-specific (e.g., anaemia, fatigue, infections), and diagnosis with risk evaluation can take time.
- Patient-level data available in EHR is currently not used systematically to regularly inform and prompt timely confirmatory testing in clinical practice.

Objective

To develop an automated, non-invasive machine learning (ML) based, risk-screening model for patients with MDS utilizing information commonly collected in routine practice.

Methods

Figure 1. Study Design using Longitudinal EHR and Claims



Modeling approach: XGBoost[1] (eXtreme Gradient Boosting) with Hazard Cox regression objective function

- Naturally handles data missingness and sparsity
- Sparsity- Aware Split Finding: Optimal default direction found by trying both directions in a split and choosing the one which proposes a maximum gain

Model Validation: 5-fold cross-validation

Model Population: 4309 MDS patients meeting inclusion criteria from ConcertAI's RWD360™ database linked with open claims data

Metric: Harrel's Concordance Index (C-Index)[2]

Results

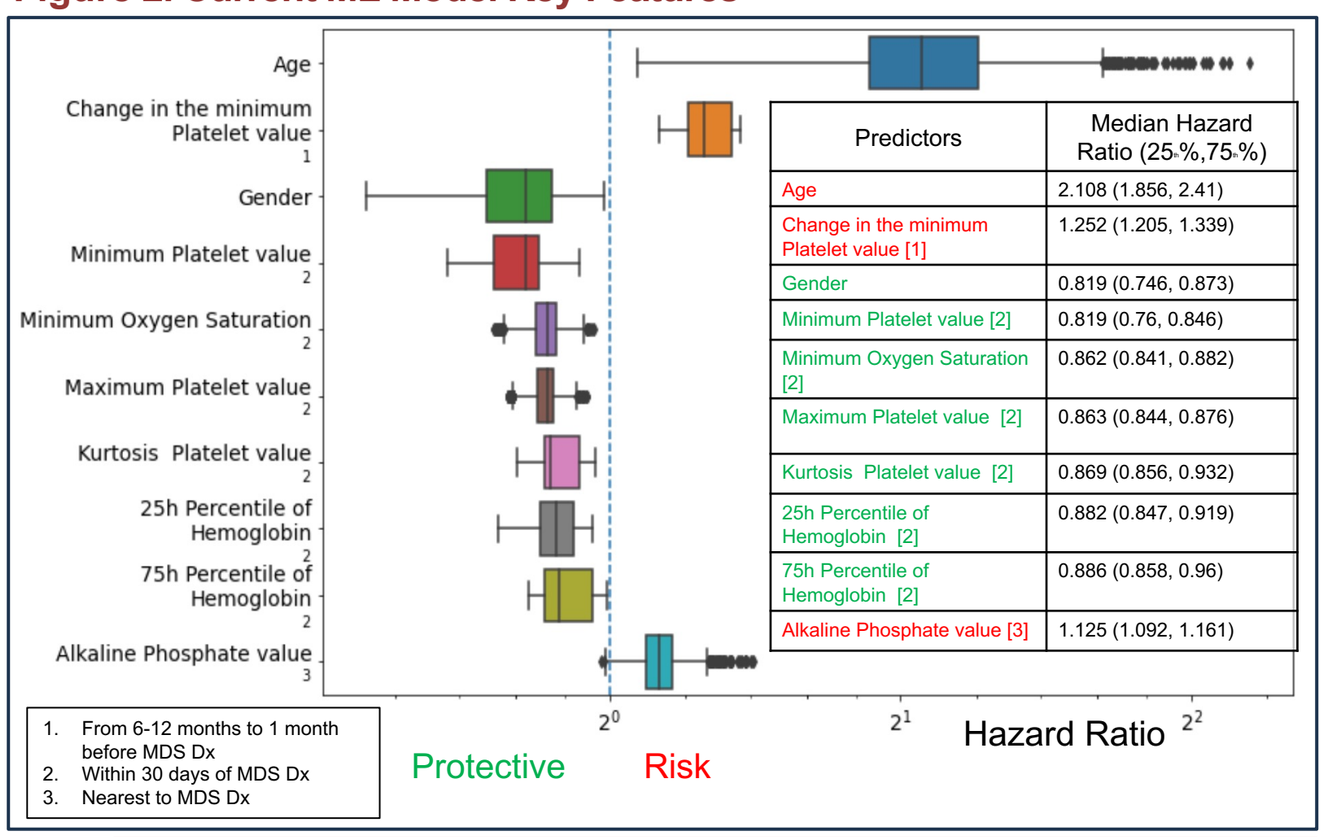
Table. ML Model Performance

MDS - Overall Survival	Proposed Model			Established Benchmarks[3,4]	
	Baseline Model no temporal features	**Primary Model** temporal features	**Streamlined Model** 4 static and 4 temporal features	IPSS-R	IPSS-M
Training C-Index	.805	.867	.855	NA	NA
Validation C-Index	.668	.699	.675	.57 to .7	.68 to .75

Training C-Index is model performance on data seen during train, validation is on data hidden from machine learning model during the training process. A 0.5 Concordance index is random chance, and 1 is a perfect model

A model streamlined for clinical simplicity using eight core features has validation C-Index of .675, only slightly behind the .699 of 130 feature primary model and comparable to established risk classification tools

Figure 2. Current ML Model Key Features



Predictors	Median Hazard Ratio (25%,75%)
Age	2.108 (1.856, 2.41)
Change in the minimum Platelet value [1]	1.252 (1.205, 1.339)
Gender	0.819 (0.746, 0.873)
Minimum Platelet value [2]	0.819 (0.76, 0.846)
Minimum Oxygen Saturation [2]	0.862 (0.841, 0.882)
Maximum Platelet value [2]	0.863 (0.844, 0.876)
Kurtosis Platelet value [2]	0.869 (0.856, 0.932)
25h Percentile of Hemoglobin [2]	0.882 (0.847, 0.919)
75h Percentile of Hemoglobin [2]	0.886 (0.858, 0.96)
Alkaline Phosphate value [3]	1.125 (1.092, 1.161)

ML model strongly selects for temporal features around the dynamics of platelet values: one year change, min and max near diagnosis, and distribution breadth as measured by kurtosis

References:
1. Chen, T, et al. *R package version 0.4-2* 1.4 2015;: 1(4#)
2. Harrell Jr FE, et al. *Statistics in medicine* 3, 1984; 3 (2#): 143-152.
3. Nazha, A, et al. *Blood* 2015; 126 (23#): 607.
4. Bernard, E, et al. *NEJM evidence* 2022; 1 (7#): EVIDoa2200008.

Acknowledgments: This study was funded by Gilead Sciences, Inc., Foster City, CA, USA.

Disclosures: **PM**, **MW**, and **CH** are employees of Gilead Sciences, Inc. **VPV**, **VP**, **DP**, **RY**, **RKD,** and **MH** are employees of ConcertAI, Inc. ConcertAI received funding from Gilead to conduct this study.

Correspondence: Cosmina.Hogea@gilead.com

Presented at the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Europe Conference; November 12-15, 2023; Copenhagen, Denmark