Use of AI/ML for prognosis of myelodysplastic syndrome using de-identified market clarity database

Khan S, Markan R, Tanwar K, Sanyal S, **Verma V**, Gaur A, Daral S, Kukreja I, Nayyar A, Roy A

RWD76

Objectives

- Myelodysplastic syndrome (MDS) is a rare blood cancer which occurs in the bone marrow with no solid tumors. Hence, conventional diagnostic assessments are not relevant for MDS. With a growing MDS patient pool, it is essential to provide efficient disease prognosis for proper treatment modalities.
- In view of the present scenario, we evaluated ML algorithms to predict MDS progression. Furthermore, to provide a more precise estimation, clinical notes were analyzed using generative Artificial Intelligence (AI).

Methods

- Optum® de-identified and consolidated Electronic Health Records (EHR) data was used for this study.
- A total of 100,164 patients, from 2017 to 2021, were identified, of which 1,434 patients were considered due to following eligibility criteria: A 12-month pre- and post-continuous enrollment, with a minimum age of 45 years, no history of MDS in the pre-index, and a minimum of 2 outpatients or 1 inpatient visit or 2 positive mentions of MDS in clinical notes
- Analyses were carried out by considering two cohorts, which include only structured EHR data, and a combination of structured EHR data with unstructured clinical notes.
- A Literature review was conducted to determine the key terms that are searched in the clinical notes.
- Generative AI was used to increase the accuracy and speed of customizing clinical notes data.
- Logistic regression, Random-Forest, and XGBoost classifier were used, considering clinical presentations, laboratory tests, and demographics as the predictors.
- To reduce the confounding effects of demographic factors, cases to controls were matched using the propensity score matching method.

Results

Table 1: Accuracy of generative AI and manual efforts for named entity recognition (NER)

Entity Type	Generative AI Accuracy	Manual effort Accuracy
Exposure to radiation	94.8%	90.5%
Benzene	96.6%	92.3%
Radioactive isotopes	98.2%	98.1%
Chemical exposure	98.2%	98.0%
Pesticide	94.2%	90.7%

Table 2: Comparison of logistic regression models based on different data sets

Model based on EHR data* only

Model	Precision	Recall	f1-score
LR	84%	79%	82%

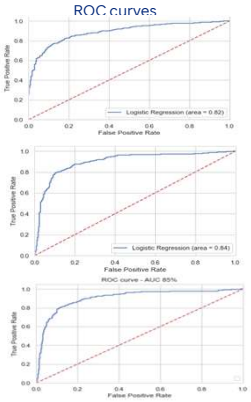
Model based on EHR structured + unstructured data using standard NLP from clinical notes

Model	Precision	Recall	f1-score
LR	86%	81%	83%

Model based on EHR structured + unstructured data using standard + customized NLP from clinical notes using generative AI

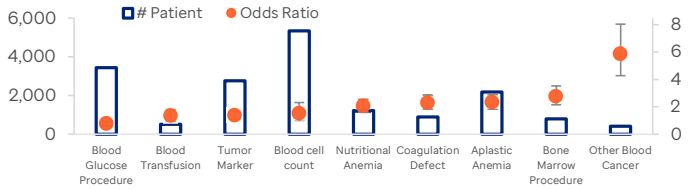
Model	Precision	Recall	f1-score
LR	87%	81%	84%

ROC curves



- Annotation and curation of customized data from clinical notes and using generative AI increased the accuracy of MDS prognosis.
- Solution developed using generative AI is ~10X efficient as compared to manual effort.
- Logistic regression was found to be the best performing model, as compared to Random-Forest and XGBoost.
- Model based on structured and unstructured EHR considering standard as well as customized NLP from clinical notes using generative AI provided the highest accuracy of 84% in comparison to the other two models (Table 2).
- Calculated odds ratios of covariates, considering 1,434 MDS patients and 4,302 patients in the reference group suggested that the highest odds ratio is represented for 'Other blood cancer'. This indicated the strongest association of 'Other blood cancer' to MDS, compared to the reference group (Figure 1).

Figure 1: Odds ratio of covariates with confidence intervals [n(MDS)= 1,434 and n(reference group)=4,302]



Conclusions

- The research findings indicate that the terminologies used in clinical notes present more precise indicators of prognostic symptoms than the structured data within EHR. This suggests that physicians have a propensity to document symptoms related to MDS and other risk factors in a patient's medical records prior to the disease's onset.
- Additionally, the utilization of generative AI has proven to be approximately ten times more efficient than manual efforts in the annotation of clinical notes.
- Early and effective prognosis of MDS can provide significant benefits to payors, providers, and pharma.

Optum