Machine learning for prospective identification of immunotherapy related adverse events

Daniel Margalski, DO, Department of Internal Medicine, Wake Forest School of Medicine, Thomas Lycan Jr., DO, Department of Hematology & Oncology, Wake Forest School of Medicine, Suraj Rajendran, Department of Biomedical Engineering, Georgia Institute of Technology, Umit Topaloglu, PhD, Department of Cancer Biology, Wake Forest School of Medicine

INTRODUCTION

The ubiquitous implementation of immunotherapy has significantly improved outcomes in the treatment of cancer patients; however, once rare adverse events from these therapies have increased in lock step. We now face an increased burden of identification on providers with limited experience in the diagnosis of irAEs. We use machine learning to develop prediction models that will aid providers in identifying patients at high risk for developing irAEs as well as for multiple downstream applications.

METHODS

We have manually extracted progress notes from 462 patients with non-small cell lung cancer treated with immunotherapy who had known irAE's, with focus on pneumonitis, colitis, and rash; the most common symptomatic irAE's. Labels were applied by clinician review to train the machine learning algorithm to identify the predictive signals at the earliest stage of recognition possible. As a standard Natural Language Processing method, we cleaned the notes to standardize punctuation, numbers and special characters. Next, we created a word embedding matrix utilizing word2vec as well as Google News Vector. Finally, we implemented a Convolutional Neural Network (CNN) on the Microsoft Azure Databricks platform. Due to class imbalance, we deployed a Synthetic Minority Over-sampling Technique algorithm as a correction. We prioritized F1 score in the analysis given the heterogeneity of the data, but will present accuracy, precision and recall as well.

RESULTS

We trained our CNN with 10 epochs resulting in an F1 of 0.428, accuracy of 0.895, precision of 0.75 and recall of 0.3 There was no significant difference in results between the word embedding matrices.

Table 1: Analysis of CNN model

Ability to predict irAE				
	Accuracy	Precision	Recall	F1
Result after 10 epochs	0.895	0.75	0.3	0.428

Conclusions

Using machine learning, we created an algorithm for irAE prediction that was accurate but lacked recall. This will serve as the foundation of implementations including the creation of a clinical decision support tool to guide focused and appropriate treatment of the unique toxicity of irAEs. Although informative as a starting point, this model had a final F1 score that was lower than expected presumably due to class imbalance of input data and the temporal nature of progress notes which limits the utility of a CNN. Future iterations of the algorithm will include supplementary documentation and implement recurrent neural networks with long short-term memory architecture to address these limitations.