MAYO CLINIC

BUILDING A NATURAL LANGUAGE PROCESSING AI TO PREDICT SUICIDE-RELATED EVENTS BASED ON PATIENT PORTAL MESSAGE DATA

ARCHIS R. BHANDARKAR MD,MS, NAMRATA ARYA BS, KELDON K. LIN BA, FREDERICK NORTH MD, MICHELLE J. DUVALL MD, **NATHANIEL E. MILLER MD, JENNIFER L. PECINA MD**

Minnesota Academy of Family Physicians
Innovation and Research Forum
March 16, 2024

DISCLOSURES

- We have no financial disclosures or conflicts of interest to disclose
- NEM and JLP are clinicians and, unfortunately, none of our NLP AI team gurus were able to attend today.
- Methodology on specific programming and techniques used is available for anyone interested:

Building a Natural Language Processing Artificial Intelligence to Predict Suicide-Related-Events Based on Patient Portal Message Data
Mayo Clin Proc Digital Health n XXX 2023;1(4):510-518 n https://doi.org/10.1016/j.mcpdig.2023.09.001
www.mcpdigitalheaorglth.

BACKGROUND

- Telemedicine utilization has increased nearly 10-fold from pre-pandemic levels and patients are increasingly using patient portals to communicate with health care teams
- US suicide rate has been generally increasing in past 15 years[1]
- Natural language processing artificial intelligence (NLP AI) can be used to detect language patterns

[1]Suicide Data and Statistics | Suicide Prevention | CDC

OBJECTIVE

- To develop an NLP AI trained on deidentified unstructured text data from patient portal messages to predict 30-day suicide-related events

METHODS

- Patient portal messages sent between Jan 1, 2013, and Oct 31, 2017, were reviewed for patients who had a suicide related event (SRE) within 30 days after sending their portal message
- SRE=death by suicide or emergency department or hospitalization for unipolar depression, suicidal ideation or suicide attempt

METHODS

- Four different machine learning algorithms used: random forest, neural network, bagged decision tree, and extreme gradient boosting
- Messages from patients with a 30-day SRE were compared with an equal number of randomly selected messages from patients without an SRE
- 80% of messages used for training the machine learning models
- 20% of messages used to compute model performance

METHODS

- Natural Language Processing "bag of words" approach used reviewing
  - Frequencies of stemmed keywords (eg 'depress' for depression, depressing, depressed)
- Metadata related to punctuation and length of message
  - Number of words in message
  - Percentage of punctuation marks that were !, ?
  - Prescence of ellipsis ("…")
  - Percentage of words in capital letters
- Sentiment analysis used to calculate average sentiment score of portal messages
  - Sentiment scores can be -1 to +1 with >0 = positive sentiment, scores <0 =negative sentiment

 |

RESULTS

- 420 messages during the study period were followed by an SRE
  - Died by suicide: 7 messages from 3 patients
  - Attempted suicide: 56 messages from 23 patients
  - Hospitalized/ED visit for depression/SI: 357 messages from 126 patients

RESULTS

- Messages with an SRE (c/w no SRE):
  - Lower average sentiment score (+0.02 vs +0.06, $p<0.001$)
  - Higher average # of words (81.2 vs 66.5, $p=0.002$)
  - Lower average proportion of !! (0.035 vs 0.055, $p=0.04$)
  - Lower average proportion of ?? (0.115 vs 0.156, $p=0.007$)
  - Higher presence of ellipses… (6.4% vs 3.1%, $p=0.02$)
  - Higher mean frequency of “help” (0.30 vs 0.16, $p<0.001$)
  - Higher mean frequency of “suicid” (0.038 vs 0, <0.001)
  - Higher mean frequency of “reall” (0.174 vs 0.074, $p<0.001$)

 |

RESULTS

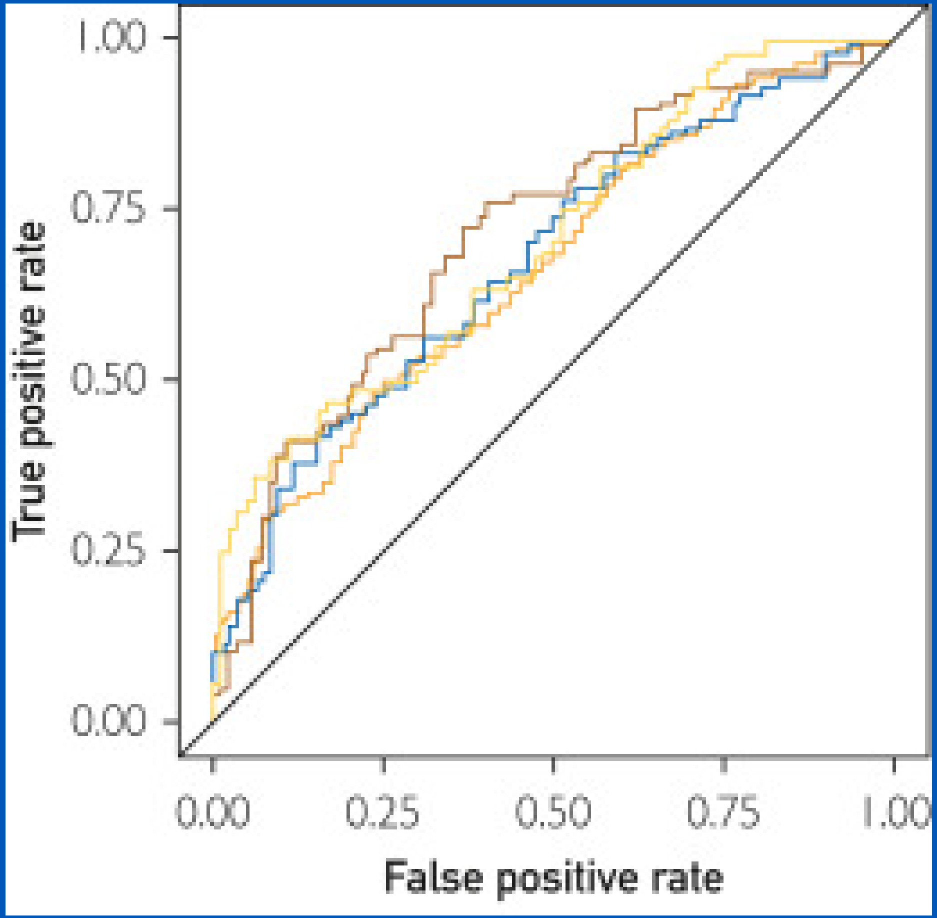
- No difference in
  - Proportion of capital letters
  - Mean frequency of “thank”
  - Mean frequency of “anxieti”

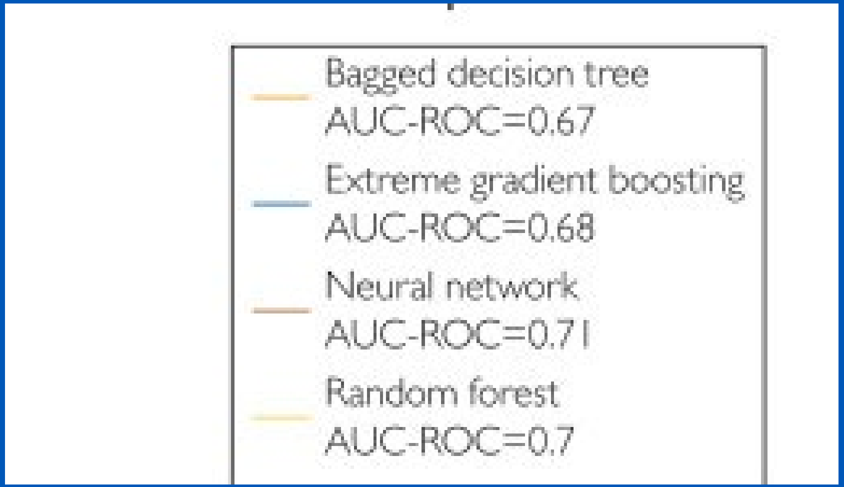
RESULTS

	SEN	SPEC	PPV	NPV	AUC-ROC
Random Forest	0.560 (0.45-0.66)	0.655 (0.55-0.75)	0.618 (0.51-0.72)	0.598 (0.50-0.69)	0.700 (0.62-0.78)
Neural Network	0.560 (0.45-0.66)	0.690 (0.59-0.78)	0.644 (0.53-0.74)	0.611 (0.51-0.70)	0.710 (0.63-0.79)
Bagged Decision Tree	0.571 (0.46-0.67)	0.619 (0.51-0.72)	0.600 (0.49-0.70)	0.591 (0.49-0.69)	0.670 (0.59-0.75)
Extreme Gradient Boosting	0.619 (0.51-0.72)	0.607 (0.50-0.70)	0.612 (0.51-0.71)	0.614 (0.51-0.71)	0.680 (0.60-0.76)

Performance of different machine learning models in predicting 30-day SREs from patient portal message features (95% CI reported in parentheses). SEN = sensitivity, SPEC = specificity, PPV = positive predictive value, NPV = negative predictive value, AUC-ROC = area under the receiver operating curve.

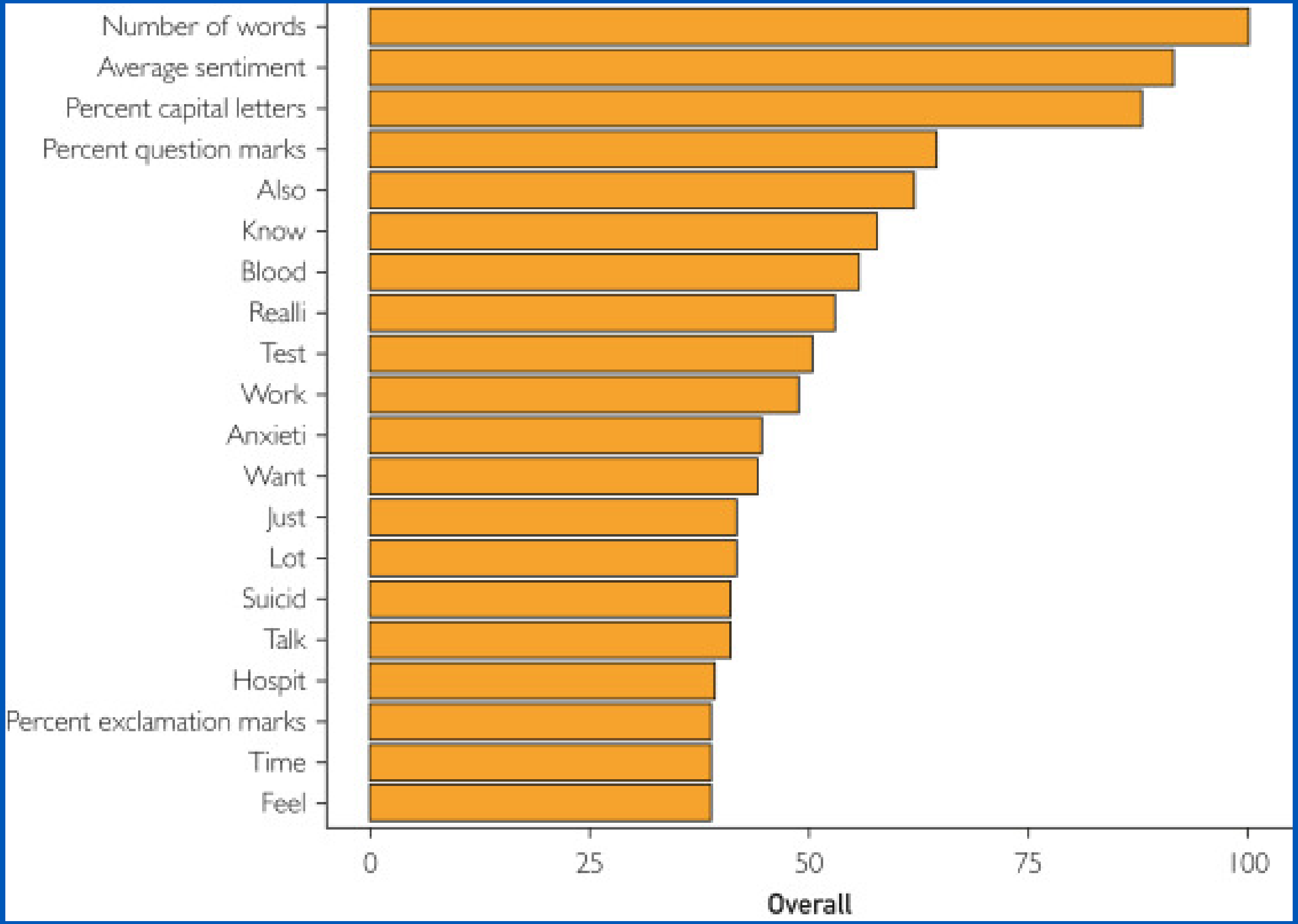
RESULTS





Comparison of receiver operating curves plotting true positive vs.
false positive rates among different machine learning algorithms on validation data

RESULTS



Explainability analysis using variable importance function to explore top features in random forest machine learning model

 |

RESULTS

Explainability analysis of top 3 features of each model showed:

- Random Forest model
  - Number of words, average sentiment score, percent capital letters
- Bagged Decision Tree model
  - Average sentiment score, proportion of capital letters, number of words
- Extreme Gradient Boosting model
  - Average sentiment score, proportion of capital letters, number of words
- Neural Network model
  - Average sentiment score, keyword “know”, keyword “also”

 |

DISCUSSION

- Early detection of SI is critical for prevention of death by suicide
- Previous studies have focused on evaluating social media posts and EHR data
- This proof-of-concept study of a limited dataset (only 420 messages) demonstrates the potential of using NLP-AI to predict SRE from unstructured text data in patient portal messages
- The 4 different models weighted features differently in terms of importance, but mean sentiment tone was in the top 3 features of all models, ie overall message tone weighted higher than individual word frequency
- Future studies could perform longitudinal analysis of changes in language patterns in portal messages over time

LIMITATIONS

- NLP models are ‘data hungry’ so different results may arise with a larger dataset
- Study did not control for ‘clustering’=the unique way of interacting with/writing portal messages of one person sending multiple messages could affect aggregate analysis
- Our study had a limited number of messages sent by patients who died by suicide

QUESTIONS & ANSWERS