LLM-Based Extraction of Psychosocial Risk Factors in Clinical Notes for Suicide-Related Risk Prediction in Cancer Patients

Maximilian Wakefield

edit

Dartmouth Health

Introduction

Cancer patients face elevated risk of suicidal ideation and self-harm, which are often documented only in free-text notes, not diagnosis codes.

Past suicide risk prediction pipelines:

Discharge Notes → LLM-Based Filter → Risk Model

Research Question & Hypothesis

Can structured EHR data and clinical notes (individually and combined) identify patterns associated with documented suicide-related events among cancer patients?

Hypothesis: Clinical notes can identify suicide-related events and provide significance in both note-only and combined machine learning models.

Data & Cohorts

Structured EHR

Structured EHR Tables from MIMIC-IV v3.1

Diagnosis codes, admissions, and demographics from the hospital record

(https://physionet.org/content/mimiciv/3.1/)

4,614 cancer patients:

- 769 **with** a documented suicide-related event
- 3,845 **without** a documented suicide-related event (matched controls)

28 features engineered from three tables:

- Admissions: hospital visit history
- Diagnosis codes: psychiatric, pain, substance-use history
- Patients: demographics (age, sex, insurance, marital status)

Benefits: Clean **clinical labels**; well-defined **structured variables**

Drawbacks: Misses **undocumented risk**, weak correlation to real-life **emotions**

Clinical Notes

Discharge summaries from MIMIC-IV v2.2

Free-text discharge summaries from the same patients

(https://physionet.org/content/mimic-iv-note/2.2/)

11,804 notes screened:

- 1,177 from cancer patients **with** a documented suicide-related event (cases notes)
- 10,627 from cancer patients **without** a documented suicide-related event (control notes)

Benefits: Mimics real-life **emotional signals** in **free text**, heterogeneous data

Drawbacks: Contains many stop words, **sparse data** (~2 notes/patient), no predefined features

Methods: LLM Filtering and Topic Modeling

LLM Filtering Pipeline

Qwen2.5-32B-Instruct

Stage 1: Extract
List every sentence describing the patient's own **negative** mood, feelings, or coping difficulty.
(In list format)

→ **Stage 2: Classify**
Is this the patient's **own** reported negative feeling?
(In Yes or No format)

→ **Stage 3: Diagnosis-Link**
Is this connected to the **cancer diagnosis** or treatment?
(In Yes, No, or Unclear format)

Figure 1: 3-stage LLM filtering pipeline overview. Extracts notes based on stage criteria (e.g., negative psychosocial factors, reactions to cancer diagnosis).

200 / 1,177 case notes kept (17.0%)	**877 / 10,627** control notes kept (8.3%)

Index date & leakage: Index date = first qualifying event (cases) or a comparable reference admission (controls). Only notes charted before the index date are used as predictors.

Topic Modeling: LDA vs. BERTopic

LDA - Full Note
Topics were:
- Organ-system dominated
- Medication dominated

BERTopic - Full Note
Topics were:
- Dosage and medication based
- Pharmaceutical language dominated

SELECTED
LDA - Evidence spans
- Coherent topics
- Clinically interpretable
- Related to suicide risk

BERTopic - Evidence spans
- Only contained 2 topics
- 1 topic absorbed most of the notes
- Not clinically interpretable

Topic count (k) selection (5-seed coherence comparison):

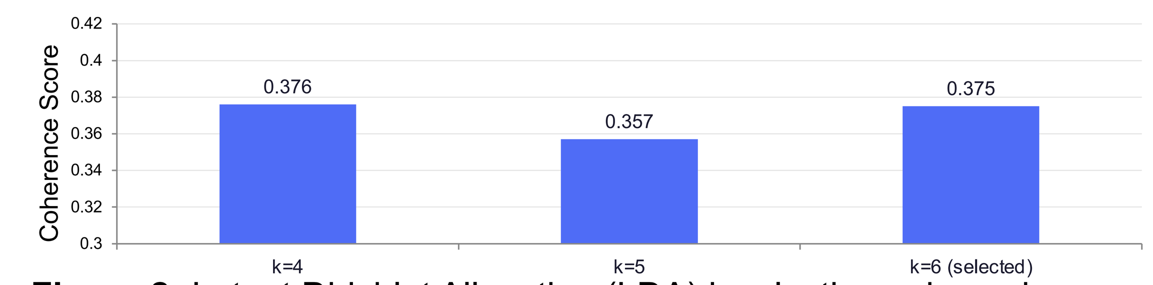

Figure 2: Latent Dirichlet Allocation (LDA) k selection using coherence scores. LDA k=6 was chosen over the marginally more stable k=4 to isolate suicidal ideation as its own theme. *Best coherence: 0.370; log-perplexity: -5.745 (train) / -9.380 (test).*

Results: LDA Topic Content & Case/Control Prevalence

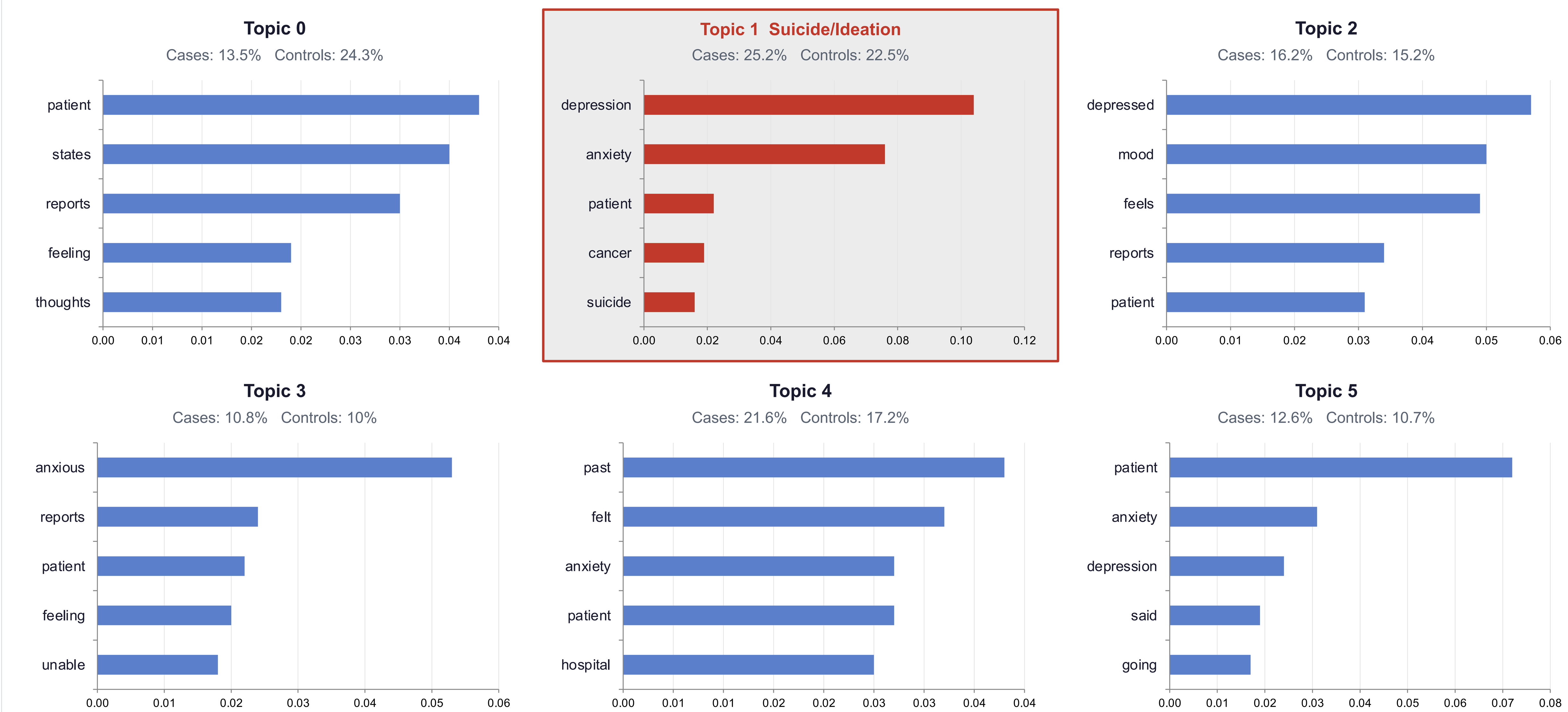

Figure 3: LDA six topics with the top five weighted words from each. The percentages show how often that topic was the main theme in a note.

Results: Predictive Value of Topic and Note Features

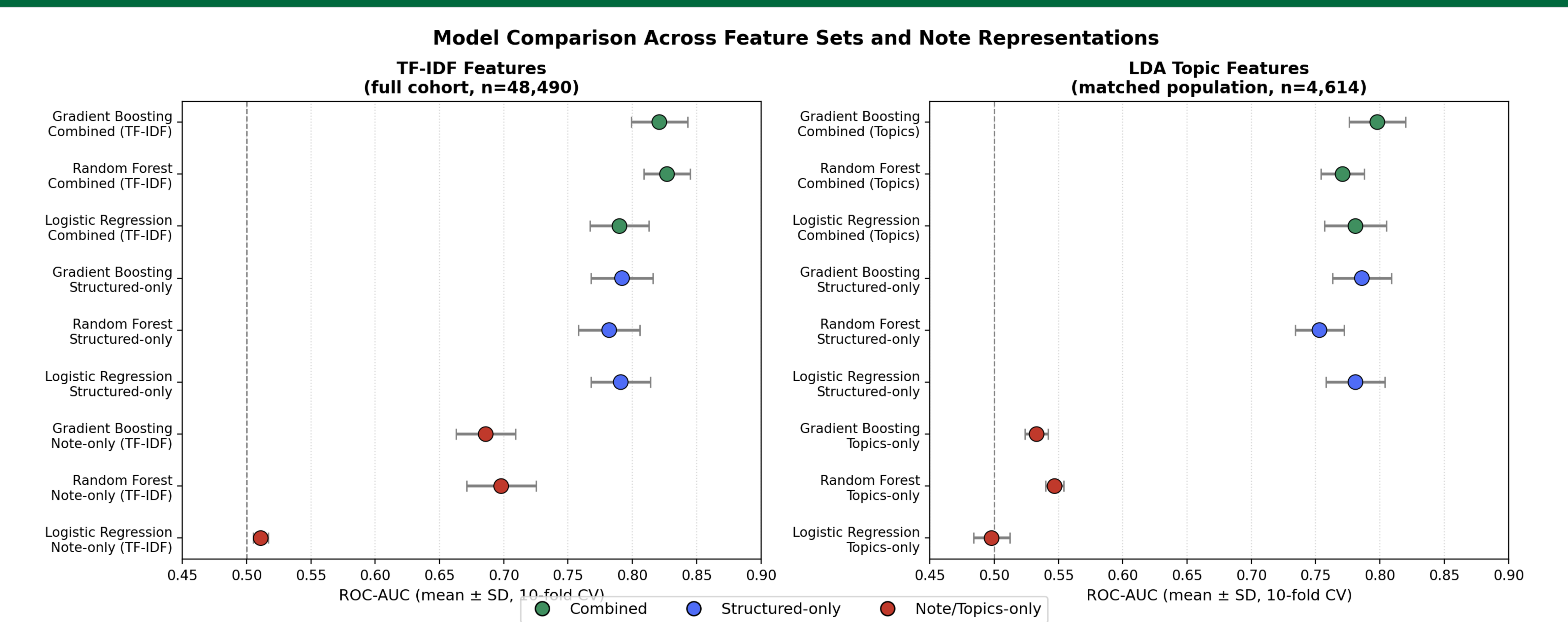

Figure 4: Model ROC-AUC for classification of cancer-related suicide risk. Error bars show mean ± SD across 10 cross-validation folds. The p-value compares Combined (structured + notes) vs. Structured-only on the same folds, isolating the added contribution of note features beyond structured data alone. Both TF-IDF and LDA topics significantly improve tree-based models (RF/GB) when added to structured data (paired t-test, $p<0.001$ for both), but not Logistic Regression.

Results: Model Accuracy Metrics

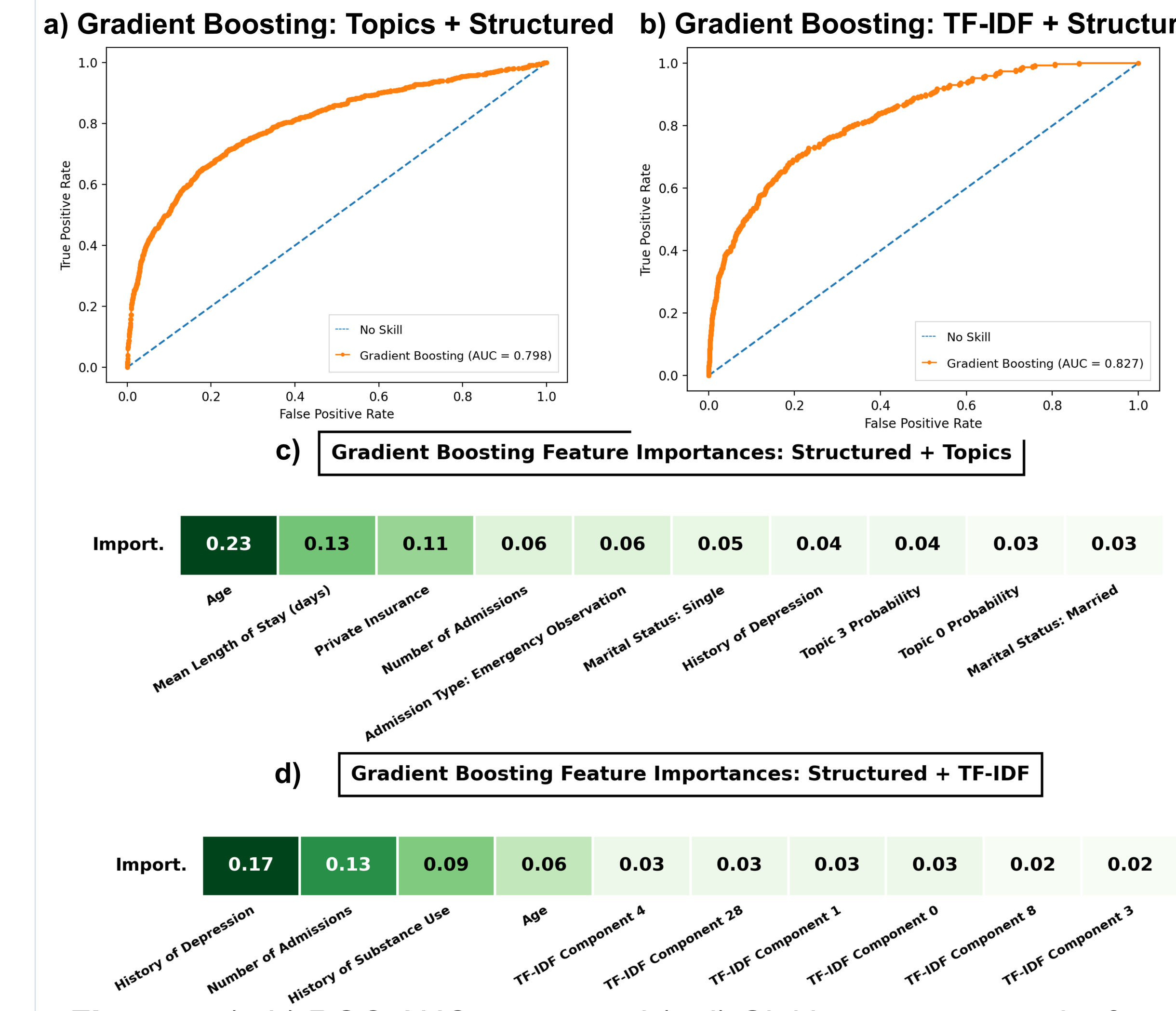

Figure 5: (a-b) ROC-AUC curves and (c-d) Gini importance graphs for classification of cancer-related suicide risk and top feature contributions for the combined Gradient Boosting models. Structured features lead both models, with note-derived features contributing psychological significance.

Longitudinal Evidence Trend

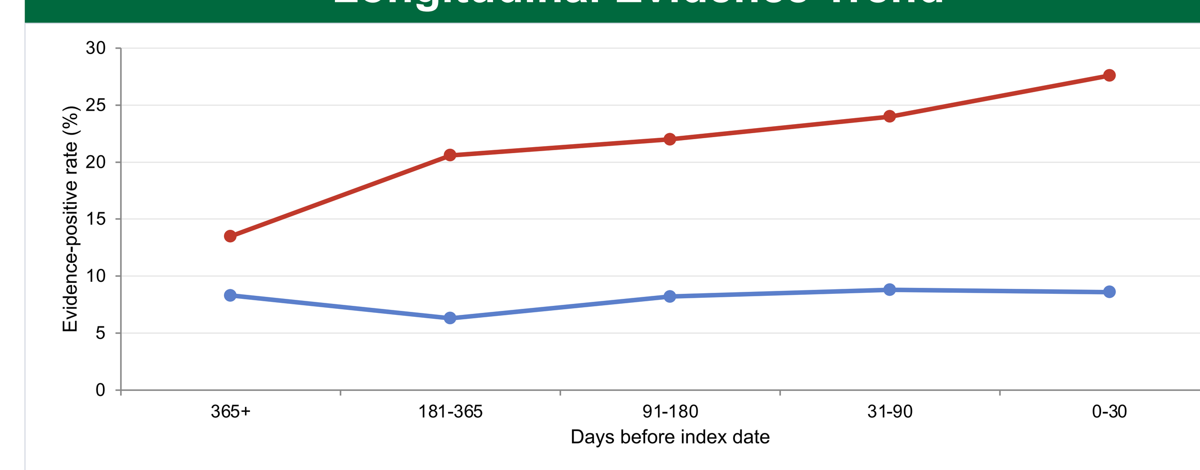

Figure 6: Longitudinal graph showing increasing case prevalence over time. Cases nearly double approaching the event; controls stay flat (interaction $p=0.0053$, $n=9{,}068$).

Conclusions

1. **Topics Help Tree Models**
Combining structured data with TF-IDF or LDA topics significantly improves reproducible Random Forest and Gradient Boosting ($p<0.001$) models, with a ROC-AUC of 0.827 and 0.798

2. **Structured Signal**
AUC 0.792 structured-only, near state-of-the-art leakage-safe accuracy metrics

3. **A Longitudinal Signal**
Evidence in notes rises toward the event in cases, producing a significant signal ($p=0.0053$)

Limitations & Future Work

Limitations
- MIMIC-IV v3.1 is **ICU/ED-only**, and has no psychotherapy notes
- Cancer-specific notes are sparse

Future Work
- Pull **broader** cancer reports for the same cohort
- Validate trajectory-based monitoring in oncology care

Acknowledgments

Dartmouth College / EDIT AI
Joshua Levy — PI & Director, EDIT AI
Louis Vaickus — Mentor, HPC & Data Infrastructure
Alos Diallo — Mentor
Benjamin Mattern, Matthew Hayden — EDIT AI Program
VA Collaborators
Monica Dimambro — Mentor
Maxwell Levis — Mentor
Data: MIMIC-IV & MIMIC-IV-Note (PhysioNet, restricted access).
Prior work: *Levis, Levy, Dimambro, Diallo, Gui, et al. LLM-based extraction of psychosocial risk factors for suicide-risk prediction in veterans. medRxiv, 2026 (preprint).*