Improving TBI Identification Through NLP-Derived Signatures from Patient-Reported Symptoms

Chenyi Ling, PhD[1], Lauren M. Diaz, MPH[2], Ellie Tyler, BS[3], Meredith M. Hughes, MA[4], Monique E. Beaudoin, PhD[5] | Abstract ID #: MHSRS-26-4170

Applied Research Laboratory for Intelligence and Security (ARLIS), University of Maryland, College Park

[1] cling@arlis.umd.edu, [2] ldiaz@arlis.umd.edu, [3] etyler@arlis.umd.edu, [4] mhughes@arlis.umd.edu, [5] mbeaudoin@arlis.umd.edu

Introduction

Traumatic brain injury (TBI) causes symptoms that overlap heavily with other conditions. Headache, memory problems, fatigue, dizziness, and mood changes appear across dozens of unrelated diseases. This overlap, spanning cognitive, somatic, vestibular, and affective domains, means TBI is frequently underrecognized or misattributed by clinicians and patients alike.

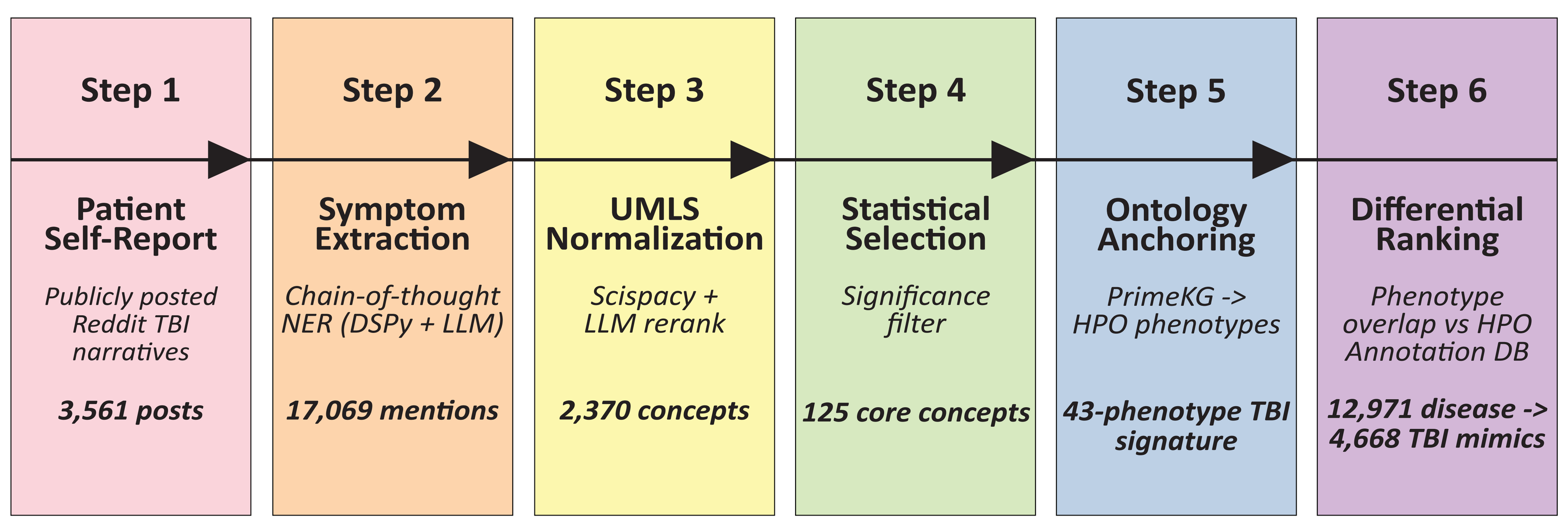
Two data sources could close this gap but are rarely combined. Patient-generated text, such as forum posts, capture rich first-person symptom descriptions, but in unstructured non-clinical language. Medical ontologies encode standardized disease-phenotype relationships, but are silent on how patients actually describe their experience. This work bridges both data sources: an automated pipeline that transforms unstructured patient-reported symptoms into a standardized phenotype signature and identifies conditions with overlapping presentations.

This work answers the question: *Given only what TBI patients write about themselves, which catalogued diseases share the most phenotypic common ground with TBI?*

Materials and Methods

We analyzed 3,561 Reddit posts written by individuals describing TBI-related symptoms. Using natural language processing (NLP) and Unified Medical Language System (UMLS) normalization, we extracted and standardized reported symptoms into a defensible phenotype signature, then ranked diseases by phenotype similarity. The full pipeline is a six-stage chain from raw narratives to a ranked differential diagnosis, followed by an independent clinical validation of the signature it produces. The full flow is summarized in the pipeline funnel (see Figure 1 below).

Figure 1. ***Pipeline funnel — From patient narratives to a disease differential.***



Step 1
Patient Self-Report
Publicly posted Reddit TBI narratives
3,561 posts
Step 2
Symptom Extraction
Chain-of-thought NER (DSPy + LLM)
17,069 mentions
Step 3
UMLS Normalization
Scispacy + LLM rerank
2,370 concepts
Step 4
Statistical Selection
Significance filter
125 core concepts
Step 5
Ontology Anchoring
PrimeKG -> HPO phenotypes
43-phenotype TBI signature
Step 6
Differential Ranking
Phenotype overlap vs HPO Annotation DB
12,971 disease -> 4,668 TBI mimics

In **step 1**, we analyzed 3,561 Reddit posts written by individuals describing TBI-related symptoms. In **step 2**, symptom mentions were extracted from patient narratives using chain-of-thought named-entity extraction. In **step 3**, mentions were normalized to UMLS concepts through a hybrid normalizer that collapses surface variants (e.g., "can't remember things," "foggy memory," "forgetfulness") onto shared concept identifiers. In **step 4**, a prevalence/co-occurrence filter reduced this set to the concepts statistically characteristic of the corpus.

In **step 5**, core concepts were anchored to the PrimeKG knowledge graph and resolved to standardized Human Phenotype Ontology (HPO) terms via curated cross-references (MRCONSO for HPO; MONDO SSSOM for disease nodes), yielding a 43-term HPO phenotype signature -- the standardized representation of "what TBI looks like" in the patient's own words (scan QR code for Figure 2). In **step 6**, this signature is compared against every disease in the HPO Annotation Database (~12,971 diseases; ~282K disease–phenotype associations), computing recall, precision, F1, Jaccard similarity, and frequency-weighted recall over the shared phenotype set.

The **final step** is independent validation. We validated the 43-phenotype signature against clinical ground truth from the Traumatic Brain Injury Model System (TBIMS) National Database (mild TBI subset, GCS category 1, n = 6,502 patients) and established diagnostic instruments (RPQ-16, DSM-5 NCD-TBI), summarized in a four-part check (S1–S4).

Results

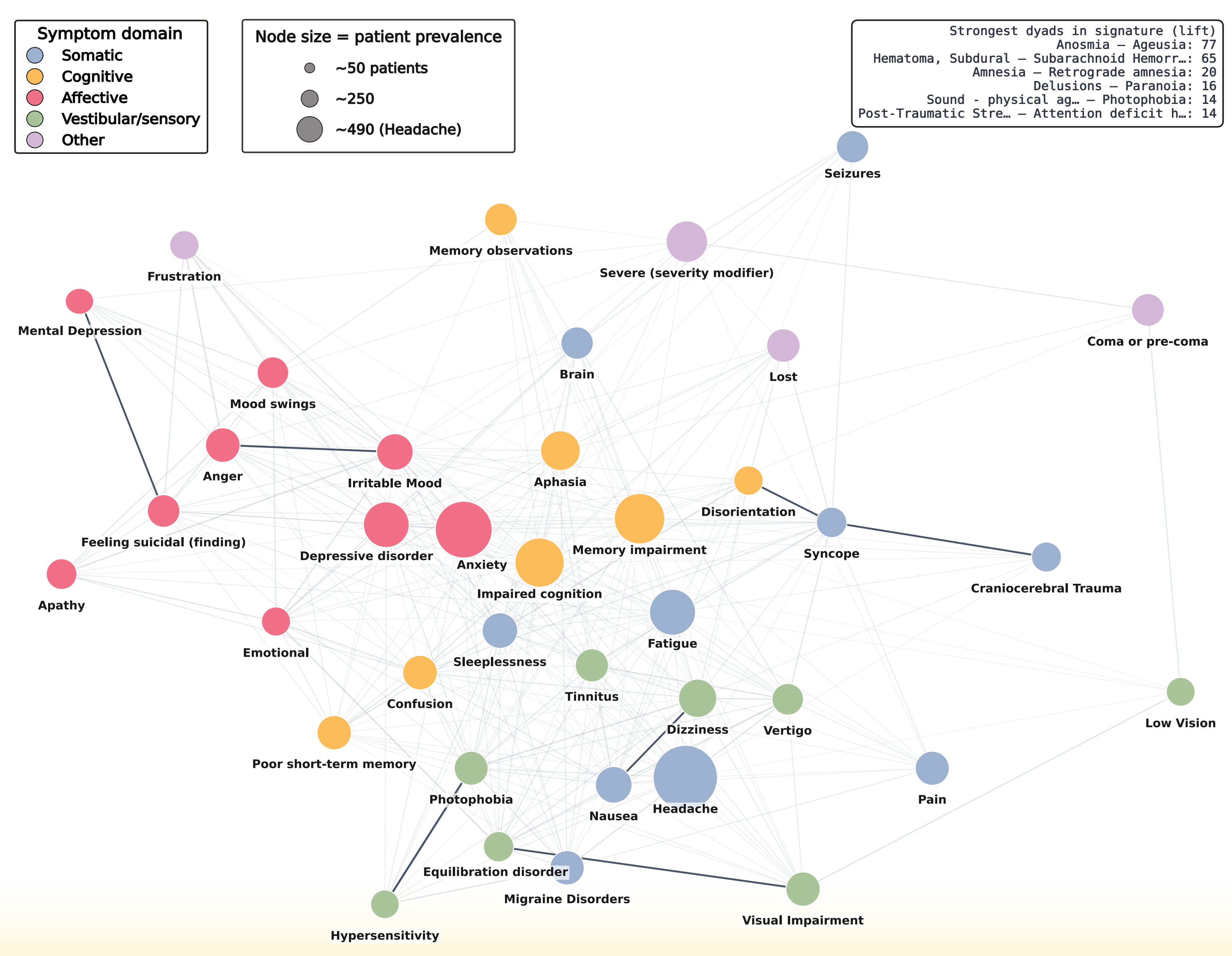
Among 12,971 conditions evaluated, 4,668 shared at least one phenotype with TBI; only 179 shared five or more, and just 11 shared 10 or more. No single condition accounted for more than 28% (12/43) of TBI phenotypes, highlighting its multisystem impact.

The TBI phenotype signature. The 43-phenotype signature recovers the recognized post-concussive profile spanning all four domains (scan QR code for Figure 3). Top anchors mirror canonical TBI presentation, for example, headache (625 mentions), memory impairment (328), cognitive impairment (301), vertigo (232), and fatigue (229). The distribution skews toward somatic and cognitive complaints with a substantial affective tail.

Clinically trustworthy signature (S1–S4). Validated against TBIMS and standard diagnostic criteria with no expert curation of the symptom list (scan QR code for Figure 4): **S1) Symptom set:** 100% of RPQ-16 symptoms recovered across all domains, and full agreement with diagnostic instruments, **S2) Symptom structure:** Unsupervised clustering reproduced all three published factor domains, with 83% concept–domain agreement, **S3) Symptom relationships:** 100% of known clinical associations reproduced; co-occurrence lift correlated with TBIMS at r = 0.79 (p = 0.019, n = 8 pairs), and **S4) Symptom hierarchy:** "Core" concepts appeared in TBIMS diagnoses at nearly twice the rate of "supportive" concepts (43% vs 24%), with 100% of expected comorbidities detected. The signature is therefore not an artifact of the Reddit corpus -- it independently matches formal clinical research on TBI.

Internal co-occurrence structure. Across 2,776 patients, the 125 core symptoms form a co-occurrence graph (683 significant edges, lift > 1) that self-organizes, unsupervised, into the same three neighborhoods recovered in S2: somatic/vestibular, cognitive, and affective (see Figure 5 below). Two association types coexist: broad, high-prevalence hubs, and tightly-bound specific dyads -- anosmia–ageusia (lift = 77), subdural–subarachnoid hemorrhage (lift = 65), and amnesia–retrograde amnesia (lift = 20) -- each clinically plausible and consistent with the S3 correlation.

Figure 5. ***Co-occurrence network — Core TBI symptoms cluster by domain and co-occur in clinically coherent dyads.***

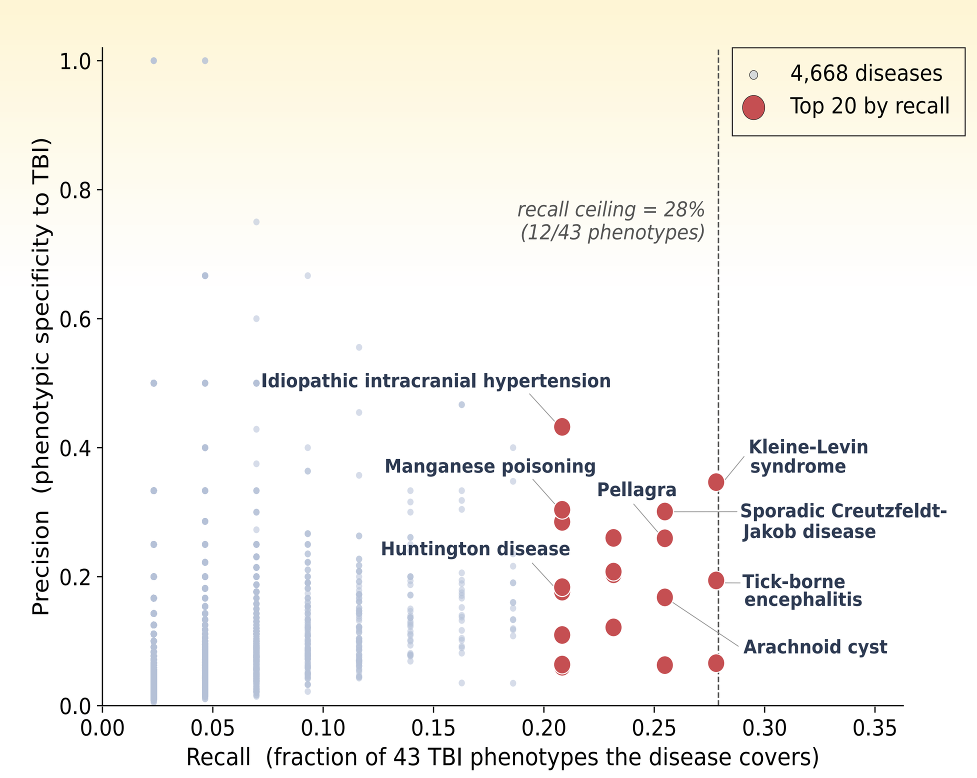


Symptom domain
Somatic
Cognitive
Affective
Vestibular/sensory
Other
Node size = patient prevalence
~50 patients
~250
~490 (Headache)
Strongest dyads in signature (lift)
Anosmia — Ageusia: 77
Hematoma, Subdural — Subarachnoid Hemorr…: 65
Amnesia — Retrograde amnesia: 20
Delusions — Paranoia: 16
Sound - physical ag… — Photophobia: 14
Post-Traumatic Stre… — Attention deficit h…: 14
Seizures
Memory observations
Severe (severity modifier)
Frustration
Mental Depression
Coma or pre-coma
Brain
Lost
Mood swings
Anger
Irritable Mood
Aphasia
Disorientation
Feeling suicidal (finding)
Depressive disorder
Anxiety
Memory impairment
Syncope
Craniocerebral Trauma
Apathy
Impaired cognition
Fatigue
Emotional
Sleeplessness
Confusion
Tinnitus
Dizziness
Vertigo
Low Vision
Poor short-term memory
Photophobia
Nausea
Headache
Pain
Equilibration disorder
Migraine Disorders
Visual Impairment
Hypersensitivity
38 UMLS concepts shown (of 125 core); edge width ∝ co-occurrence lift; dark edges = strongest dyads

Cont.

Disease differential ranking. By recall, top conditions are clinically coherent (scan QR code for Figure 6): tick-borne encephalitis and Kleine-Levin syndrome (12/43 each), followed by sporadic CJD (11/43), arachnoid cyst (11/43), pellagra (11/43), manganese poisoning, Huntington disease, and IIH -- all capable of a TBI-like presentation.

Figure 8. ***Precision–recall scatter — No single disease covers TBI's full breadth.***



4,668 diseases
Top 20 by recall
recall ceiling = 28%
(12/43 phenotypes)
Idiopathic intracranial hypertension
Manganese poisoning
Pellagra
Huntington disease
Kleine-Levin syndrome
Sporadic Creutzfeldt-Jakob disease
Tick-borne encephalitis
Arachnoid cyst
Precision (phenotypic specificity to TBI)
Recall (fraction of 43 TBI phenotypes the disease covers)
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35

By F1, leaders shift toward conditions whose own profile is dominated by TBI-type phenotypes (scan QR code for Figure 7): Kleine-Levin syndrome (F1 = 0.308), IIH (precision = 0.429, F1 = 0.281), sporadic CJD (0.275), and pellagra (0.256). Precision–recall separates high-precision mimics (IIH, Mal de débarquement, glioblastoma, CJD) from high-recall mimics (tick-borne encephalitis, Kleine-Levin) -- two distinct flavors of "looks like TBI" (see Figure 8 to the left).

Among the top 100 mimics, six categories dominated (scan QR for Figure 9): neurodegenerative (11), CNS infection (9), toxic/ metabolic encephalopathy (8), autoimmune/ inflammatory (5), intracranial mass/pressure (4), and vestibular/ sleep/behavioral (3) -- the same disease families that diffusely affect cognition, mood, and somatic function like TBI.

Discussion

Summary. This approach turns unstructured patient narratives into standardized phenotype signatures, offering a scalable way to improve TBI recognition and inform clinical and public awareness. The method surfaced recognized TBI mimics without clinician input; IIH exemplifies this with the highest precision (0.429). No disease exceeds ~28% recall, reflecting TBI's simultaneous impact across cognitive, somatic, vestibular, and affective systems. Sparse annotation of common conditions (e.g., migraine, ranked ~250th) means rankings reflect phenotype profile similarity, not clinical likelihood. The S1-S4 validation grounds the differential ranking established clinical knowledge, not corpus idiosyncrasies.

Limitations. 1.) Sensitivity/specificity against confirmed diagnoses were not measured; rankings are hypothesis-generating, not a decision rule, 2) sparsely annotated common conditions are systematically under-ranked; absolute rank should not be read as clinical likelihood, 3) the Reddit corpus is predominantly mild TBI- and concussion-related and self-selected, 4) nearly a third (30.6%) of symptom mentions were unmappable, thus some signal is dropped, and 5) results depend on the HPO Annotation Database snapshot and will shift as annotations are updated.

Conclusions. Patient self-report, processed end-to-end with no expert curation of the symptom list, can be transformed into a standardized HPO phenotype signature and used to rank ~13,000 diseases into a clinically coherent differential. The signature independently reproduces established TBI clinical knowledge, and the rankings recover recognized TBI mimics, concentrated among inflammatory, neurodegenerative, and toxic encephalopathies. The approach generalizes to any condition for which patient-generated symptom text exists, offering a scalable, ontology-grounded route from unstructured narrative to a structured differential.

Acknowledgments

This work was supported by the Defense Health Agency (DHA) Research & Engineering (R&E) under contract HT942524C0036. Special thanks to our sponsor for their continued support.