

---

**Review**

---

# Natural Language Processing Applied to Psychiatric Clinical Notes: Scoping Review

---

Shuying Rao<sup>1,2,3,4</sup>, BEng; Xi'ang Chen<sup>1,2,3,4</sup>, MEng; Guifeng Deng<sup>1,2,3,4</sup>, BEng; Junyi Xie<sup>1,3,4</sup>, BEng; Tiecheng Jiang<sup>5</sup>, MEng; Tao Li<sup>1,3,4</sup>, PhD; Yaoyun Zhang<sup>6</sup>, PhD; Haiteng Jiang<sup>1,3,4</sup>, PhD

<sup>1</sup>School of Brain Science and Brain Medicine, and Liangzhu Laboratory, Zhejiang University School of Medicine, Affiliated Mental Health Center and Hangzhou Seventh People's Hospital, Hangzhou, Zhejiang, China

<sup>2</sup>College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, Zhejiang, China

<sup>3</sup>MOE Frontier Science Center for Brain Science and Brain-machine Integration, Zhejiang University, State Key Laboratory of Brain-machine Intelligence, Hangzhou, Zhejiang, China

<sup>4</sup>Zhejiang Key Laboratory of Clinical and Basic Research for Psychiatric Diseases, Hangzhou, Zhejiang, China

<sup>5</sup>Department of Biomedical Engineering, Shenyang University of Technology, Shenyang, Liaoning, China

<sup>6</sup>The University of Texas at Dallas, Richardson, TX, United States

## Corresponding Author:

Haiteng Jiang, PhD

School of Brain Science and Brain Medicine, and Liangzhu Laboratory

Zhejiang University School of Medicine, Affiliated Mental Health Center and Hangzhou Seventh People's Hospital

305 Tianmushan Road, Xihu District

Hangzhou, Zhejiang 310058

China

Phone: 86 0571-87071107

Email: [h.jiang@zju.edu.cn](mailto:h.jiang@zju.edu.cn)

---

## Abstract

**Background:** Psychiatric clinical notes in electronic health records (EHRs) provide rich longitudinal information that can support clinical decision-making. Using historical medical data can enable earlier identification of mental illness, better characterization of disease trajectories, and more personalized treatment planning. Natural language processing (NLP) transforms these unstructured notes into analyzable representations for research and care.

**Objective:** This study aims to systematically summarize NLP methodologies for psychiatric clinical notes, compare major modeling paradigms and application areas, and highlight emerging large language model (LLM) trends, key challenges, and future research directions.

**Methods:** Following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines, a literature search was conducted for articles on NLP methods based on psychiatric clinical notes published from January 2021 to December 2025 in Ovid MEDLINE, Ovid EMBASE, PubMed, Scopus, Web of Science, the ACM Digital Library, and ScienceDirect. This scoping review analyzed NLP methods applied to psychiatric clinical notes, focusing on major trends, identifying suitable features for traditional machine learning (ML)-based models, applications of pretrained language models (PLMs), and key challenges. Approaches were categorized as rule-based, traditional ML, hybrid, deep learning (DL), and LLM-based methods across information extraction and text classification tasks.

**Results:** In total, 101 studies were eligible for inclusion. Rule-based methods (n=36) and hybrid approaches (n=34) remained the most widely used techniques, largely favored for their interpretability in handling nuanced, subjective clinical notes. These were followed by DL (n=15), traditional ML (n=10), and LLM-based approaches (n=6). Traditional ML studies relied heavily on engineered features, which could be grouped into 5 broad categories: domain knowledge features, lexical and statistical features, vector-based semantic features, emotion-related features, and temporal features. PLMs improved performance mainly through domain adaptation and task-specific fine-tuning, enhancing the handling of psychiatric language, medical terminology, and clinical note structure. LLM-based studies, although still limited in number, indicated a growing shift toward generative and reasoning-based applications.

**Conclusions:** Hybrid NLP approaches remain dominant, combining domain rules with ML for extraction and classification. DL approaches continue to advance, with domain adaptation supporting medical terminology and clinical semantics. LLMs

may further automate complex workflows via zero-shot capabilities and reasoning, alongside growing interest in temporal modeling and multimodal integration. Key future needs include improved generalizability across institutions, privacy protection, and careful attention to ethical implications in clinical deployment.

*JMIR Med Inform* 2026;14:e91249; doi: [10.2196/91249](https://doi.org/10.2196/91249)

**Keywords:** psychiatric clinical notes; electronic health records; natural language processing; large language models; information extraction

## Introduction

Mental illness, also known as psychiatric disorders, is a prevalent problem worldwide and continues to be one of the most serious public health issues [1]. There are many different types of mental illness, including depression, suicidal ideation, bipolar disorder, autism spectrum disorder (ASD), anxiety disorders, schizophrenia, etc. All kinds of mental illness negatively affect an individual's physical health and well-being, and the COVID-19 epidemic has further exacerbated this problem [2]. According to recent statistics, millions of people worldwide experience one or more psychological disorders [1]. Therefore, it is of great significance to study the development trajectory and potential mechanisms of mental illness based on scientific and objective measurement methods, and to promote the development of early diagnosis, personalized medicine, and precision treatment strategies for mental illness.

Language plays a central role in mental health, serving as a medium for expressing symptoms, delivering therapy, and assessing clinical conditions. Traditionally, language analysis in psychiatry has relied on expert opinions, clinical ratings, and manual methods, which are often subjective, incomplete, or prone to inaccuracies. Automated language analysis offers a transformative opportunity to shift from subjective clinical judgment to “measurement-based care” [3], enabling robust, quantitative, and scalable tracking of language variables. This advancement has the potential to revolutionize psychiatric practice and research.

Electronic health records (EHRs) are a rich source of health care data and have been widely used to record patient medical histories [4]. Psychiatric EHRs often include narrative clinical notes that contain valuable information for advancing clinical research and health care [5]. However, much of this clinical information remains locked in unstructured text [6], posing challenges for systematic analysis. In the field of psychiatry, standardized measures of patients are used inconsistently and infrequently in clinical practice. Due to the heterogeneity and complexity of clinical notes, along with its diverse application scopes, the application of natural language processing (NLP) in clinical notes still needs to be further explored.

The rise of statistical NLP in the 1990s [7] and recent advances in deep learning (DL) technology [8,9] have influenced the methods used in clinical notes analysis. A pivotal turning point in this trajectory was the introduction of the Transformer architecture, which enabled a new modeling paradigm: pretrained language models (PLMs), such as Bidirectional Encoder Representations from Transformers

(BERT) [10]. PLMs are first trained on large, general-purpose text corpora via self-supervised objectives, then adapted to specific downstream tasks through fine-tuning — substantially reducing dependence on large manually annotated datasets. With the rapid expansion of computational resources, PLMs were subsequently scaled to billions of parameters, giving rise to large language models (LLMs), such as GPT and Gemini, which demonstrate superior performance across a wide range of NLP tasks, including text classification, entity recognition, summarization, sentiment analysis, and text generation [11]. Clinical applications for clinical notes, such as summarization and information extraction [12,13], have been widely used to assist in disease diagnosis, including suicide screening [14], depression identification [15], and mental state prediction [16].

Several recent reviews have explored various aspects of NLP in mental health, each with distinct focuses. Some reviews are highly specialized, concentrating on specific aspects such as mental illness detection [17] or intervention tools [18]. Other reviews do not incorporate the recent breakthroughs, such as the development and application of attention mechanisms, transformers, and pretrained LLMs. Reviews such as Le Glaz et al [19] primarily focus on traditional NLP and machine learning (ML) techniques applied across mental health domains. Meanwhile, Jin et al [20] provide a high-level overview of LLM applications and performance metrics without diving into the specific methodological challenges of working with clinical notes. Given these gaps, this scoping review aims to provide a comprehensive and methodologically detailed examination of NLP methods applied specifically to psychiatric clinical notes. We focus on the latest trends, tools, and challenges in using NLP for mental illness research based on clinical notes.

This scoping review aimed to address the following research questions:

1. What are the major NLP trends and methods for clinical notes analysis in psychiatric disorders?
2. In traditional ML-based models, which features of clinical notes are suitable for extraction for downstream research on mental illness?
3. How to apply PLMs to improve the performance of text-based models in the field of mental illness?
4. What are the main challenges and future directions for NLP in clinical notes of psychiatric disorders?

## Methods

### *Design and Protocol Registration*

This scoping review was conducted in accordance with the Joanna Briggs Institute guidance for scoping reviews and reported in compliance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [21-23]. A formal protocol was not prospectively registered. The review protocol was retrospectively registered in the Open Science Framework [24]. The completed PRISMA-ScR checklist can be found in [Checklist 1](#).

Our search included any document published from January 2021 to December 2025. We selected 2021 as the starting year to position this review as an update to prior broad mental health NLP reviews and to focus on the most recent methodological phase of the field.

### *Information Sources and Search Strategy*

A systematic electronic search was developed with an experienced librarian and implemented across 7 databases: Ovid MEDLINE, Ovid EMBASE, PubMed, Scopus, Web of Science, the ACM Digital Library, and ScienceDirect. The search was first performed on January 16, 2024, and then updated with a second search on December 15, 2025.

The search strategy combined controlled vocabulary terms, where available, and free-text terms covering 3 core concepts: (1) psychiatric or mental disorders, (2) clinical notes or electronic health records, and (3) NLP or text mining. Database-specific syntax, subject headings, field tags, and adjacency operators were adapted for each source while preserving the same conceptual structure across databases. The full search strategies for all databases are reported in [Multimedia Appendix 1](#).

All retrieved citations were exported to Zotero. Duplicate records were removed in 2 stages: (1) automated deduplication using bibliographic fields such as title, author, year, journal, and DOI or PMID, followed by (2) manual review of potential duplicate pairs.

### *Inclusion and Exclusion Criteria*

We included English-language full-text journal articles and full conference or workshop papers that reported completed studies using NLP methods applied to English psychiatric clinical notes, and that made a methodological contribution. Studies using quantitative, qualitative, or mixed-method designs were eligible if they described completed research.

We excluded abstract-only publications, study protocols, review articles, editorials, and papers without sufficient full-text methodological detail. We also excluded studies in which the primary text source was not psychiatric clinical notes; examples included interviews, speech or voice data, social media posts, or general clinical text unrelated to mental health. Studies were also excluded if they did not apply NLP methods or if their application was not specific to psychia-

try or mental illness (eg, deidentification or redundancy removal).

### *Study Selection and Screening Process*

Initial screening of the titles and abstracts was conducted independently by 2 reviewers either (SR and TJ) or (GD and JX). The records were divided between 2 reviewer pairs (SR with TJ, and GD with JX), and each record was assessed by 1 reviewer pair. At this stage, each article was categorized into one of the following groups: (1) fully met the inclusion criteria, (2) did not analyze clinical data, (3) did not use NLP methods, (4) did not focus on mental illness, and (5) had unclear eligibility for inclusion. To maximize sensitivity, records were advanced to full-text review whenever eligibility could not be determined from the title and abstract alone.

Full-text eligibility assessment followed the same procedure, with each report independently assessed by 2 reviewers within 1 reviewer pair. At this stage, all inclusion criteria had to be met for a study to be retained in the review. Disagreements at either stage were first resolved through discussion within the reviewer pair; if consensus could not be reached, a third reviewer (XC) adjudicated the final decision.

### *Data Collection*

The final data collection form used for peer-reviewed articles is shown in Table S1 in [Multimedia Appendix 2](#). The information for each study included data sources, sample information, regions, NLP tasks, applied NLP methods, application domain, and psychiatric categories. These articles were divided into 5 parts (each part included 20 or 21 papers). A total of 5 researchers (SR, XC, GD, JX, and TJ) independently extracted data from each part of the papers. To ensure the accuracy of data extraction, we conducted a double-check process. The researchers cross-checked each other's work, reviewing the data extraction process.

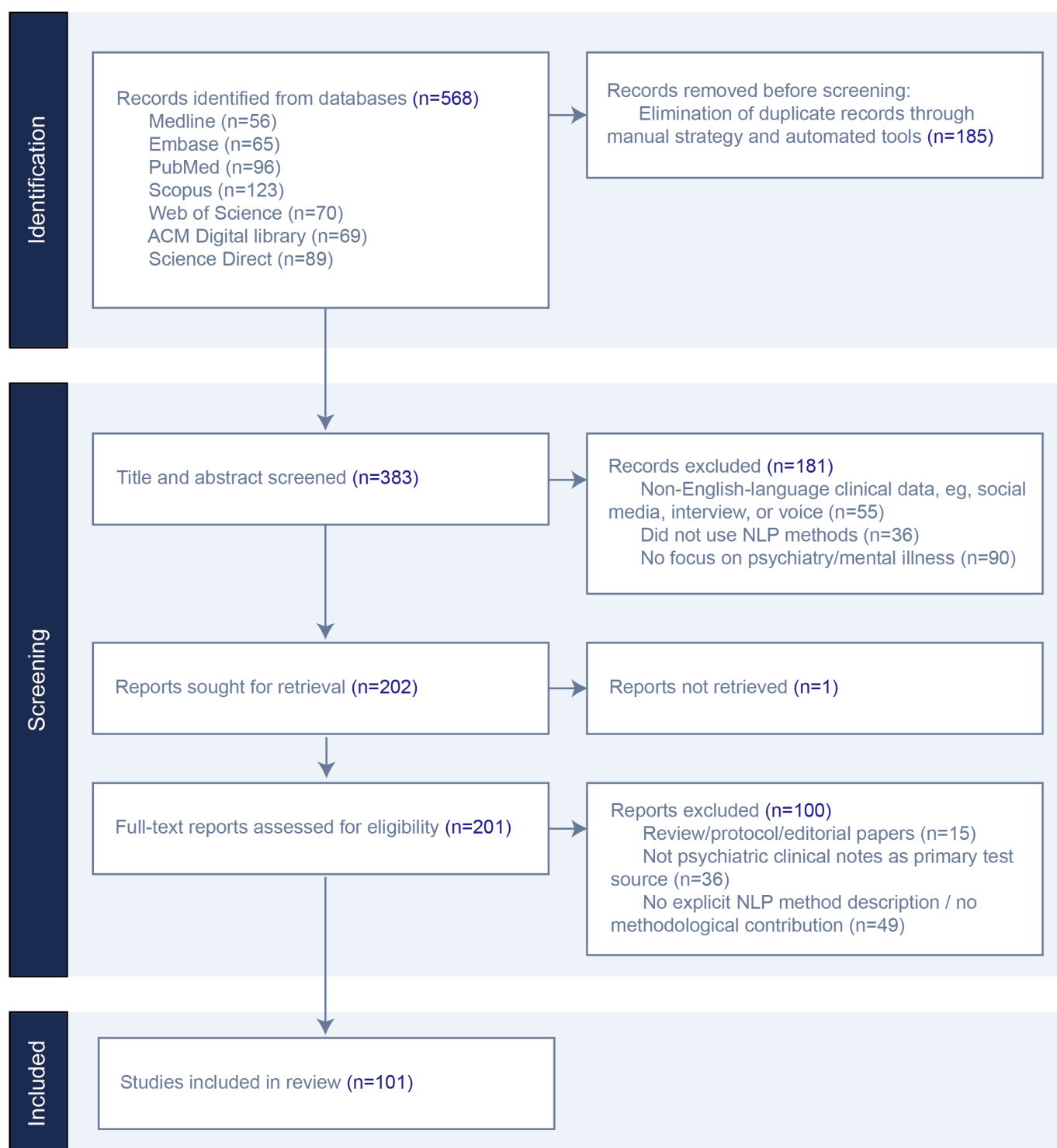
For this scoping review, we developed a categorization framework based on NLP methodology approach applied for psychiatric clinical notes: rule-based, traditional ML, hybrid, DL, and LLM-based methods. Definitions of the key terms used throughout this review are provided in [Multimedia Appendix 3](#).

## Results

### *Overview*

The initial search yielded 383 records after deduplication. Following title and abstract screening, 181 records were excluded, leaving 202 reports for full-text retrieval. Of these, 1 report could not be retrieved, and the remaining 201 were assessed for eligibility. Full-text screening resulted in the exclusion of a further 100 articles, with reasons detailed in [Figure 1](#). Ultimately, 101 studies [25-125] were included in the scoping review, as shown in [Multimedia Appendix 4](#). The results of the search and the study inclusion process are presented in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart in [Figure 1](#) [126].

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of search history. EHR: electronic health records; NLP: natural language processing.

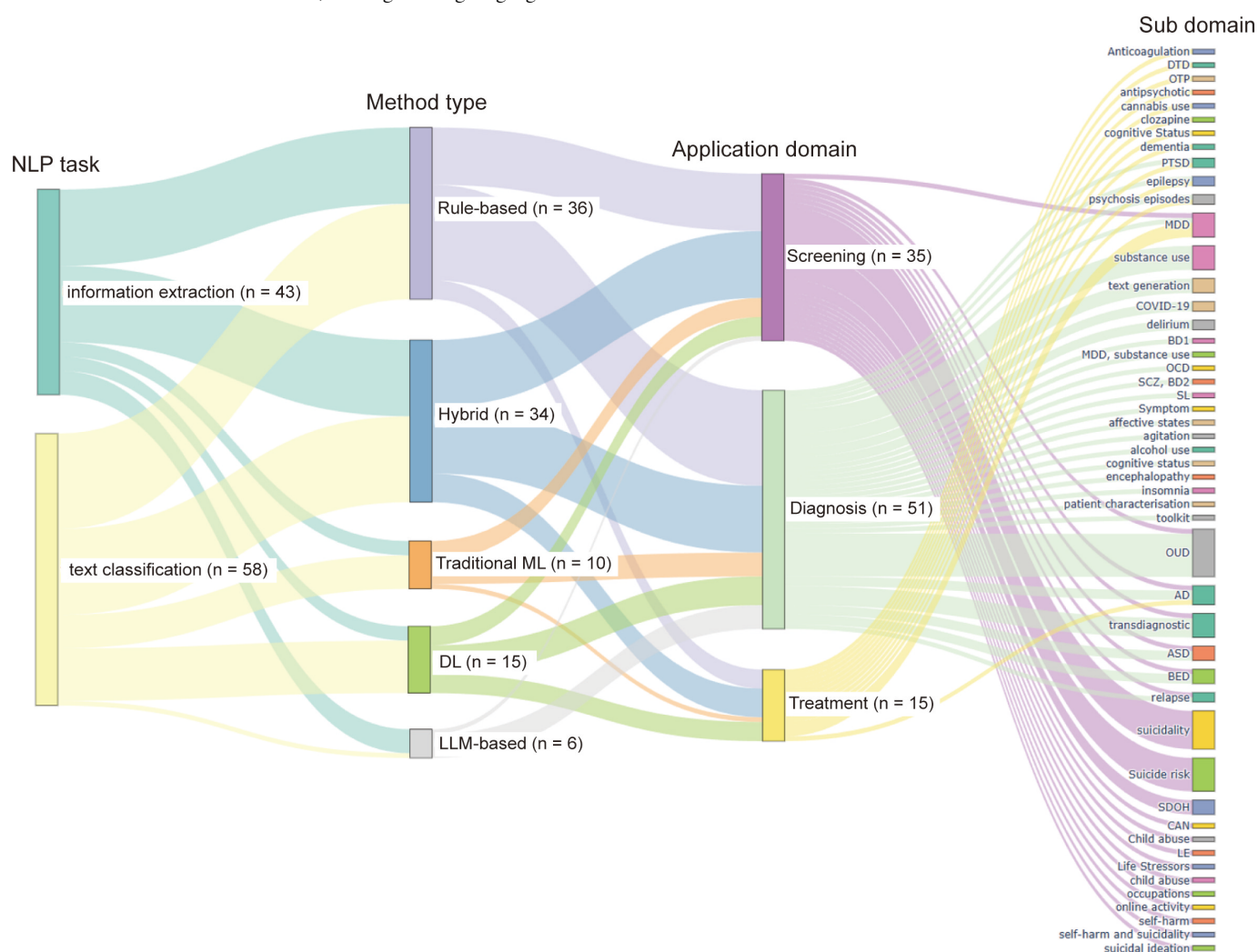


Across the 101 included studies [25-125], the identified tasks were broadly grouped into information extraction (IE) and text classification (TC), implemented using rule-based, traditional ML, hybrid, DL, and LLM-based approaches. Figure 2 presents a Sankey diagram [127] illustrating the relationships among NLP tasks, methods, and application domains in mental health. Specifically, the included studies were organized into three major application domains: screening, diagnosis, and treatment. Screening-related studies mainly focused on suicidal ideation, suicide

risk, crisis prediction, and related risk detection; diagnostic applications involved symptom identification, disease classification, phenotype characterization, and comorbidity recognition; and treatment-related studies addressed drug information extraction, treatment quality assessment, and outcome prediction. Together, these pathways highlight the interdisciplinary nature of applying NLP to psychological assessment and intervention by demonstrating how various computational techniques interconnect across different domains.



**Figure 2.** Sankey diagram of natural language processing (NLP) tasks, methods, and clinical applications across the 101 included studies. The diagram is read from left to right, linking NLP task categories (information extraction and text classification) to method types (rule-based, hybrid, traditional machine learning (ML), deep learning (DL), and large language model [LLM]–based approaches), then to the 3 main clinical application domains (screening, diagnosis, and treatment), and finally to specific psychiatric subdomains. The width of each flow is proportional to the number of studies following that pathway, thereby showing how specific NLP tasks and methods connect to downstream clinical applications. AD: Alzheimer disease; ASD: autism spectrum disorder; BD1: behavioral disturbance; BD2: bipolar disorders; BED: binge-eating disorder; CAN: child abuse and neglect; CVD: cardiovascular disease; DTD: difficult-to-treat depression; LE: life events; MDD: major depressive disorder; OCD: obsessive-compulsive disorder; OTP: opioid treatment program; OUD: opioid use disorder; PTSD: posttraumatic stress disorder; SCZ: schizophrenia; SDOH: social determinants of health; SL: stigmatizing language.



The following subsections detail the resources (including data sources and knowledge bases) and a systematic review of modeling approaches ranging from rule-based and traditional ML to hybrid, DL, and the latest LLM-based approaches. Finally, we address model evaluation metrics and performance. Although many techniques overlap with general clinical NLP, the distinctive characteristics of psychiatric notes and mental health care challenges require specialized resources, methods, and considerations.

## Resources

Clinical notes in the reviewed studies were primarily sourced from medical institutions, with access typically requiring data use applications and institutional review board approval due to privacy concerns. Among the 101 included studies, data were drawn from 42 medical institutions and only one public database, the MIMIC-III dataset [128], underscoring the limited availability of openly accessible psychiatric EHR

resources. Specialized psychiatric data sources, particularly the South London and Maudsley NHS Foundation Trust [129] and the US Department of Veterans Affairs (VA) [130], were among the most frequently used, appearing in 24 and 9 studies, respectively. In addition, Rush University Medical Center's Substance Use Intervention Team (SUIT) [131] program contributed data to five studies on opioid addiction. Overall, the heavy reliance on institution-specific proprietary databases indicates substantial fragmentation in the current resource landscape, which may hinder reproducibility, external validation, and cross-institutional collaboration. A detailed distribution of data sources is provided in [Multimedia Appendix 5](#).

The Observational Medical Outcomes Partnership (OMOP) common data model is an open standard designed to standardize observational data, enabling efficient and reliable analyses [132]. In [25,26], the OMOP common data model

was used to harmonize clinical data, including diagnostic codes, demographics, and clinical notes. Similarly, the Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR) standard facilitates health care data exchange using modern web technologies, organizing data into modular “resources” accessible via RESTful APIs [133]. Afshar et al [27] used HL7 data standards to transfer EHR data to the cloud, supporting the deployment of a real-time NLP CDS tool for opioid misuse screening. However, the adoption of these standards in mental health remains limited due to challenges such as customization requirements, integration complexity, and privacy concerns. Future advancements in data-sharing frameworks and public dataset availability may enhance research in this field.

Standard terminology sets are foundational for NLP in psychiatric clinical notes by providing unified concepts and formats that support interoperability across institutions and

systems. Commonly used terminologies include the Unified Medical Language System (UMLS) [134], Systematized Nomenclature of Medicine–Clinical Terms (SNOMED-CT) [135], *International Classification of Diseases, Ninth Revision (ICD-9)* and *International Classification of Diseases, Tenth Revision (ICD-10)* [136], and RxNorm [137] for drug information. Psychiatric clinical notes analysis also faces distinct needs—complex diagnostic criteria, behavioral assessments, subjective language, strict confidentiality, and longitudinal monitoring—so domain-specific resources are often essential. The *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V)* provides standardized diagnostic criteria for mental disorders [138], while the Addiction Behaviors Checklist (ABC) supports structured assessment of addictive behaviors [139]. Links to these terminology sets are provided in Table 1.

**Table 1.** An overview of the dictionaries and knowledge bases used for extracting information (data) from clinical notes.

| Dictionary or knowledge base | Description                                                                               | Examples |
|------------------------------|-------------------------------------------------------------------------------------------|----------|
| UMLS <sup>a</sup> [134]      | Biomedical thesaurus organized by concept and it links similar names for the same concept | [28]     |
| SNOMED-CT <sup>b</sup> [135] | A globally accepted medical terminology system                                            | [29]     |
| ICD <sup>c</sup> [136]       | Knowledge on the extent, causes, and consequences of human disease and death worldwide    | [30]     |
| DSM <sup>d</sup> [138]       | Standardized classification and criteria for diagnosing mental disorders                  | [31]     |
| RxNorm [137]                 | A standardized nomenclature for clinical drugs                                            | [32]     |
| ABC <sup>e</sup> [139]       | A scale commonly used to assess an individual’s addictive behavior                        | [33]     |

<sup>a</sup>UMLS: Unified Medical Language System.

<sup>b</sup>SNOMED-CT: Systematized Nomenclature of Medicine–Clinical Terms.

<sup>c</sup>ICD: *International Classification of Diseases*.

<sup>d</sup>DSM: *Diagnostic and Statistical Manual of Mental Disorders*.

<sup>e</sup>ABC: Addiction Behaviors Checklist.

Rule-Based Approaches

Rule-based approaches use predefined rules and keyword-based features to identify patterns in text [140], offering transparency, traceability, and cost-effectiveness by leveraging domain-specific knowledge without requiring large annotated datasets [141]. These methods often rely on clinical standards, guidelines, curated dictionaries, and knowledge bases, which can be easily updated and adapted [142]. Initial keyword lists are typically derived from expert domain knowledge or standard terminologies like UMLS, SNOMED-CT, and *ICD-9* and *ICD-10*, enabling comprehensive coverage of medical concepts. Rule-based NLP pipelines map free-text clinical notes to standardized ontologies (eg, UMLS) through medical concept normalization, identifying terms related to drugs, diagnoses, symptoms, and clinical measurements while expanding searches to related concept groups.

Clinical notes often contain nonstandard language, abbreviations, misspellings, and diverse expressions for the same symptom. Rule-based approaches refine terminology dictionaries using regular expression (RegEx) patterns and iterative feedback from domain experts [34,35]. They also identify modifiers—such as emotions (negation, affirmation), descriptive attributes (severity, duration), and annotation sections (medical history, assessment)—to contextualize

concepts and exclude irrelevant mentions [32,36]. Negation detection, often using the NegEx algorithm [143], is a key application. Rule-based methods can also track symptom trajectories, as demonstrated by Young et al [37], who classified and monitored dynamic behavioral phenotypes in ICU patients over time.

However, relying solely on dictionary-based concept recognition cannot capture all essential information in clinical texts. To capture comprehensive information, custom rule sets are often developed, as detailed in [144]. Several NLP systems facilitate rule-based development, including MedLEE [145], MetaMap [146], clinical text analysis and knowledge extraction system (cTAKES) [147], Clinical Language Annotation, Modeling, and Processing Toolkit (CLAMP) [148], and Biomedical Information Collection and Understanding System (BioMedICUS) [149], which are widely used for named entity recognition (NER) and IE in clinical and biomedical research [150].

Traditional Machine Learning Approaches

Traditional ML refers to a family of statistical and mathematical algorithms that learn patterns from data to make predictions or decisions, but typically require human-guided feature engineering rather than automatically extracting hierarchical

representations through multilayered neural networks. In TC tasks, traditional ML models such as Conditional Random Fields, support vector machines, structured support vector machines, logistic regression, Bayesian models, and random forests are commonly used. These models learn patterns from input data and labeled outputs without explicit programming [151], with their development process involving data preprocessing, feature extraction, modeling, optimization, and evaluation. Unlike DL, traditional ML requires significant human intervention for feature engineering. As summarized in Table 2, the most suitable features fall into five broad categories: domain knowledge features, lexical and statistical features, vector-based semantic features, emotion-related features, and temporal features. These features capture

different aspects of psychiatric clinical notes that are relevant to downstream research. Domain knowledge and lexicon-based features are useful for identifying explicit symptom descriptions and clinically salient terminology; semantic vector representations help normalize variation in narrative expression; emotion-related features, extracted using tools like ABSApp [152], are particularly relevant for affective states and suicide risk [38]; and temporal features are important for modeling symptom progression [39], instability, and longitudinal risk. Together, these findings suggest that traditional ML is especially suited to settings where clinically interpretable features can be engineered from notes to support tasks such as phenotyping, subgroup discovery, outcome prediction, and risk stratification.

**Table 2.** An overview of features used in traditional machine learning-based models.

| Features                         | Description                                                                                                     | Examples |
|----------------------------------|-----------------------------------------------------------------------------------------------------------------|----------|
| Domain knowledge features        |                                                                                                                 |          |
| UMLS <sup>a</sup>                | UMLS is a set of key terminology, coding standards, and associated resources related to biomedical information. | [28,29]  |
| Lexical and statistical features |                                                                                                                 |          |
| BoW <sup>b</sup>                 | The simplest form of text representation using numbers of vocabularies.                                         | [40]     |
| n-gram                           | N-gram is a contiguous sequence of n words.                                                                     | [41]     |
| TF-IDF <sup>c</sup>              | TF-IDF reflects the importance of the word in the document.                                                     | [31,42]  |
| Vector-based semantic features   |                                                                                                                 |          |
| Word embedding                   | The vector-based representation of words. Examples: word2vec, GloVe.                                            | [43]     |
| Emotion-related features         |                                                                                                                 |          |
| Sentiment scores                 | Determining the sentiment polarity of texts (positive, negative, or neutral)                                    | [38,44]  |
| Temporal features                |                                                                                                                 |          |
| Time features                    | Focusing on the time-related features, like time interval.                                                      | [39]     |

<sup>a</sup>UMLS: unified medical language system  
<sup>b</sup>BoW: Bag of Words  
<sup>c</sup>TF-IDF: term frequency–inverse document frequency

Most ML research in this field relies on supervised learning, which requires high-quality labeled data for effective model training. However, data labeling is time-consuming and challenging. Unsupervised learning methods, such as clustering [153] and latent Dirichlet allocation (LDA) topic modeling [45], can extract useful patterns without labeled data and may complement supervised classifiers [46]. Clustering methods, such as non-negative matrix factorization (NMF), can differentiate patient subgroups based on text features. Zhao et al [31] used TF-IDF to normalize ASD terms in clinical texts, revealing distinct ASD subgroups through clustering. Topic modeling, an unsupervised technique for identifying thematic patterns in text [154], represents topics as probability distributions over words. LDA topic modeling has successfully identified insomnia factors within military health systems [45] and transformed unstructured “chief complaints” into quantitative symptom clusters (eg, somatic or cognitive distress) [47]. Additionally, dynamic topic modeling (LDASeq) models how word distributions evolve over time by inducing conditional dependence between sequential intervals. By tracing the longitudinal prominence of themes such as “Suicide” or “Medication,” Levis et al [48] provide a mechanism for monitoring clinical “lability” and time-sensitive risk fluctuations in high-risk populations, offering insights into the

dynamic nature of suicide risk that traditional static variables cannot capture.

Hybrid Approaches

The hybrid approach integrates rule-based and ML methods within a single system. In terminal hybrid models, rule-based systems perform feature extraction, producing structured outputs that serve as inputs for ML models. Rule-based methods, leveraging dictionary lookup and pattern recognition, facilitate medical concept normalization. Feature engineering then vectorizes extracted data for ML applications. Ontologies further enhance NLP by capturing related concepts and contextual nuances, such as negation and speculation in clinical notes. For instance, phrases like “patient denies memory loss” or “possible onset of dementia” are accurately processed. IY Oh et al [49] applied this approach to extract clinical phenotypes from unstructured data, which was then used to train ML models for predicting Alzheimer disease progression and identifying modifiable risk factors.

Hybrid approaches also enhance ML-driven lexicon expansion and disambiguation. Developed open-source NLP software such as NimbleMiner [155] allows users to mine clinical texts to quickly discover large synonym vocabularies

containing abbreviations and spelling errors based on word embedding models. For concept normalization, the Medical Concept Annotation Toolkit (MedCAT) [156] uses self-supervised embeddings to disambiguate candidates detected via dictionaries, accurately mapping mentions to the SNOMED-CT ontology even amidst linguistic noise.

Foundational infrastructures like the General Architecture for Text Engineering [157] facilitate hybrid NLP by enabling the integration of rule-based transducers, such as Java Annotation Patterns Engine, with diverse ML resources. A key application is TextHunter [158], which powers the CRIS-CODE project [159] by combining SVM for sentence classification with the rule-based ConText algorithm. By automating the detection of negation, temporality, and subject, TextHunter has successfully extracted over 40

mental health symptoms with a 90% median precision, supporting transdiagnostic research [50,51] and dynamic psychosis risk prediction [52]. Moreover, the CLARK [160] toolkit democratizes these hybrid methods for non-experts by allowing users to define clinical features through RegEx, which are then transformed into feature vectors for downstream ML classifiers. CLARK has demonstrated robust performance in identifying depression and substance use disorder diagnoses within medically complex populations [53]. Lately, VIEWER [54] enhanced mental health care by integrating hybrid NLP into visual analytics, providing a longitudinal, data-driven perspective of patient journeys. An overview of several prominent clinical NLP toolkits in psychiatry based on these hybrid approaches is provided in Table 3.

**Table 3.** An overview of clinical natural language processing toolkits in psychiatry based on hybrid approaches.

| Toolkit                 | Description                                                                  | Rule-based method                                              | ML <sup>a</sup> method                                       | Examples       |
|-------------------------|------------------------------------------------------------------------------|----------------------------------------------------------------|--------------------------------------------------------------|----------------|
| GATE <sup>b</sup> [157] | A modular infrastructure for the full lifecycle of text analytics            | JAPE <sup>c</sup> grammars for pattern matching and gazetteers | Pluggable models (SVM <sup>d</sup> , Weka, DL) <sup>e</sup>  | [55]           |
| TextHunter [158]        | A suite designed for psychiatric concept extraction and model building       | ConText algorithm for negation and subject detection           | SVM for high-precision sentence classification               | [51,52]        |
| CLARK [160]             | A graphical interface for computable phenotyping by clinical researchers     | User-defined RegEx for feature selection                       | Standard ML classifiers (RF <sup>f</sup> , SVM, Naïve Bayes) | [53]           |
| NimbleMiner [155]       | A system for rapid lexicon discovery and mining through word similarity      | High-precision regex search using enriched lexicons            | Skip-gram word embeddings for finding semantic synonyms      | [56]           |
| MedCAT [156]            | A toolkit for automated concept linking and disambiguation in clinical notes | Dictionary-based concept candidate detection                   | Self-supervised embeddings (Word2Vec) for disambiguation     | [57]           |
| VIEWER [54]             | An interactive visual analytics tool for point-of-care decision support      | Pattern matching for symptom and intervention detection        | Distributed pipelines for automated clinical data extraction | Not available. |

<sup>a</sup>ML: machine learning.  
<sup>b</sup>GATE: General Architecture for Text Engineering.  
<sup>c</sup>JAPE: Java Annotation Patterns Engine.  
<sup>d</sup>SVM: support vector machine.  
<sup>e</sup>DL: deep learning.  
<sup>f</sup>RF: random forest.

Deep Learning Approaches

Traditional ML models rely on feature engineering, whereas DL frameworks automatically capture meaningful features without manual intervention [161]. DL, a subfield of ML, learns hierarchical representations through neural networks such as recurrent neural networks (RNNs), convolutional neural networks (CNNs), and transformers.

DL frameworks typically consist of an embedding layer and a classification layer [162]. Embedding techniques range from word embeddings to contextual encoders (eg, BERT [10] and ALBERT [163]). In classification layers, CNNs use convolutional filters to capture spatial relationships and pooling layers to reduce computational complexity. In [58], a CNN-based multilabel classifier was developed to predict alcohol abuse, opioid abuse, and non-opioid substance abuse. This model demonstrated higher sensitivity than single-label classifiers, suggesting that integrating multiple substance abuse screenings into a single model can enhance clinical decision support and reduce alarm fatigue.

Beyond custom architectures, researchers have leveraged modular NLP frameworks to address specific psychiatric

tasks with high data efficiency. By fine-tuning ScispaCy [164] to extract “health status” keywords from outpatient notes, Verter et al [59] achieved high precision (>92%) using minimal manual annotations. In addition, Med7 was a DL model originally trained on general physical health data [128], and in [165], it was further fine-tuned on the UK-CRIS database to extract granular pharmacological details (including dosages and titration schedules), thus enabling the comprehensive characterization of treatment resistance in a cohort of over 28,000 patients.

Recent advances in NLP have introduced PLMs — a class of deep neural models pretrained on large text corpora and adapted to specific tasks through fine-tuning. These PLMs, predominantly built on the Transformer architecture, demonstrate strong generalization across tasks and domains [10]. For instance, Ford et al [60] fine-tuned the BERT model for NER to identify 5 key concepts (diagnosis, medication, dosage, signs or symptoms, and substance use), coupled with a contextual classification model to determine the “Status” (eg, has, had, and does not have) and “Experiencer” (eg, patient vs. family member) of the extracted entities. This methodology was integrated into the Akrivia



Health database [61] framework to process millions of patient records. Domain-specific adaptations such as BioBERT [166], MentalBERT [167], and Bio\_ClinicalBERT [168] have also proven highly effective in clinical settings [62,169]. Building on these discriminative strengths, Xie et al [63] further integrated the generative capabilities of the text-to-text transfer transformer (T5) model [170] to capture complex epilepsy outcomes. T5 was used to transform unstructured notes into precise, structured data regarding seizure frequency and dates, effectively revealing the remitting-relapsing dynamics of epilepsy across large-scale EHRs.

LLM-Based Approaches

While PLMs transformed the field by facilitating the fine-tuning of broad representations for targeted clinical tasks, the emergence of generative LLMs marks a distinct evolutionary leap [171]. In this review, LLMs refer specifically to PLMs scaled to billions of parameters and built predominantly on decoder-only Transformer architectures. This massive scaling unlocks “emergent” abilities, most notably zero-shot chain-of-thought (CoT) reasoning [172] and in-context learning [173], allowing them to simulate complex clinical decision-making and generate coherent patient narratives without weight updates or extensive task-specific training data [174]. For example, Leng et al [64] demonstrated that GPT-4o, using a “summary of summaries” hierarchical approach, could identify stages of cognitive impairment with an expert concordance rate (Kappa=0.95) that far surpassed BERT-based benchmarks. Moreover, iterative prompt refinement has enabled models like GPT-4 to produce discharge summaries that blinded psychiatry specialists rated as superior to those written by residents [65]. This reasoning capability also facilitates the generation of high-fidelity synthetic data, a crucial development for fields constrained by stringent privacy regulations. Warner et al [66] leveraged 2 local instances of Llama 3.3 (70B) to act as “Interviewer” and

“Patient” through CoT-driven interaction; they generated complex synthetic patient profiles that provide a viable alternative to sensitive psychiatric clinical notes for model training.

To bridge the gap between general reasoning and domain-specific precision, researchers have increasingly focused on fine-tuning. Techniques such as low-rank adaptation (LoRA) [175] allow models to master the nuanced lexicon of psychiatry while operating within computational and privacy constraints. Notably, Shukla et al [67] illustrated that smaller, open-source models (eg, Llama-3-8B) fine-tuned via LoRA can outperform massive proprietary models like GPT-4o in specialized tasks, including note proofreading and substance use identification. This trend toward localized optimization is further exemplified by the work of Krishnamoorthy et al [68], who used a fine-tuned Llama 3.3 (70B) to translate complex discharge summaries into patient-friendly language, thereby enhancing health literacy and engagement.

To situate these recent developments within the broader language-model landscape of psychiatric NLP, Table 4 provides a comparative overview of the major language model families applied to psychiatric clinical notes, including both earlier encoder-based PLMs and recent generative LLMs. It demonstrates how foundational open-source encoders are typically fine-tuned for granular extraction, whereas modern scalable models leverage advanced prompting and parameter-efficient tuning to tackle complex reasoning, summarization, and synthetic generation. It also suggests a degree of functional differentiation across model types: general-purpose proprietary LLMs showed their clearest advantages in longitudinal summarization, reasoning-oriented staging, and patient profile generation, whereas domain-adapted or locally fine-tuned open models were often more competitive for tightly bounded extraction, standardization, and institution-specific workflows.

Table 4. Overview of language model families applied to psychiatric clinical notes.

| Studies       | Model family   | Representative models                                                                                                                                                               | Adaptation strategy                                                                                                                                                                                                                  |
|---------------|----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Open-source   |                |                                                                                                                                                                                     |                                                                                                                                                                                                                                      |
| [5,10,60,69]  | BERT series    | <ul style="list-style-type: none"><li>• BERT<sup>a</sup></li><li>• ClinicalBERT</li><li>• BlueBERT,</li><li>• BioClinicalBERT</li><li>• MentalBERT</li><li>• DementiaBERT</li></ul> | <ul style="list-style-type: none"><li>• Continued pretraining on clinical or psychiatric corpora</li><li>• Supervised fine-tuning (eg, for Named Entity Recognition)</li><li>• Embedding regularization (eg, triplet-loss)</li></ul> |
| [7,11,66,68]  | LLaMa series   | <ul style="list-style-type: none"><li>• Llama 3.1</li><li>• Llama 3.2</li><li>• Llama 3.3</li></ul>                                                                                 | <ul style="list-style-type: none"><li>• Parameter-efficient tuning (eg, low-rank adaptation)</li><li>• Local instruction tuning (with institution-specific note corpora),</li><li>• Locally deployed multiagent prompting</li></ul>  |
| [70]          | FLAN-T5 series | <ul style="list-style-type: none"><li>• FLAN-T5-XL</li></ul>                                                                                                                        | <ul style="list-style-type: none"><li>• Zero-shot instruction followed by category-specific fine-tuning</li><li>• Synthetic-example augmentation</li><li>• Question-answering style prompting for note-level extraction</li></ul>    |
| Closed-source |                |                                                                                                                                                                                     |                                                                                                                                                                                                                                      |
| [4,11,64]     | GPT series     | <ul style="list-style-type: none"><li>• ChatGPT-4</li><li>• GPT-4o</li></ul>                                                                                                        | <ul style="list-style-type: none"><li>• Prompt-based inference (multistep, confidence-aware)</li><li>• Integration with retrieval system</li><li>• Clinician-in-the-loop review</li></ul>                                            |

<sup>a</sup>BERT: Bidirectional Encoder Representations from Transformers.

Model Evaluation

In IE and TC tasks, performance is typically evaluated using a confusion matrix (contingency table) to derive error rates from true positives, false positives, false negatives, and true negatives. Common metrics include sensitivity, recall, specificity, precision (PPV), NPV, and  $F_1$ -score. For probabilistic outputs, threshold-independent measures such as AUC and PRAUC are also reported. Most ML studies adopt a hold-out design (training, validation, and test), while cross-validation (CV) estimates predictive error by repeatedly training on subsets and testing on the remainder.

LLM evaluation often follows a multidimensional framework that combines quantitative metrics with expert review. Human assessment—eg, blinded preference tests, distinguishability studies, and Likert-scale rubrics—focuses on clinical utility, coherence, and correction effort. Automated metrics (eg, BERTScore [176], METEOR [177], readability indices [178]) complement human judgment by quantifying semantic fidelity and text quality. Safety and reliability are commonly examined through hallucination and prompt-adherence error analysis, and, when generating synthetic data, statistical comparisons with real-world demographic distributions. Interannotator consistency is

frequently measured using Cohen kappa, reflecting variability due to fatigue, interpretation differences, and annotator expertise.

Table 5 summarizes representative model evaluation results across typical IE and TC tasks, including domain-specific evaluation strategies: (1) temporal evaluation, reflecting symptom progression over time (eg, Garriga et al [39], area under the receiver operating characteristic curve=0.865 for mental health crisis prediction); (2) multilabel evaluation, addressing comorbidity (eg, Afshar et al [58], area under the receiver operating characteristic curve=0.88 for alcohol misuse and 0.94 for opioid misuse); (3) severity-aware metrics, particularly relevant for suicide risk where false negatives are high-cost; (4) clinician agreement, benchmarking model outputs against clinician judgments (eg, Leng et al [44], 95% consistency for CI stages classification); (5) cross-domain generalization, comparing performance across conditions or health care systems (eg, Cliffe et al [71]); and (6) unsupervised evaluation, using clinical interpretability of discovered patterns when labels are limited (eg, Andrew et al [46] applying topic modeling and clustering to opioid-related cohorts).

Table 5. Results of model evaluation on information extraction and text classification tasks.

| Categories and study         | Primary aim                                                                                                            | Subjects or dataset                                                 | Performance or main findings                                                                                                         |
|------------------------------|------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|
| Rule-based                   |                                                                                                                        |                                                                     |                                                                                                                                      |
| [72]                         | Identify neuropsychiatric symptom domains following COVID-19 hospitalization                                           | 6619 patients from 6 Eastern Massachusetts hospitals                | The most commonly-documented symptom domains were fatigue (13.4%), mood and anxiety symptoms (11.2%), and impaired cognition (8.0%)  |
| [71]                         | Characterize each eating disorder patient's suicidality profile                                                        | 1126 and 420 patients at WCM <sup>a</sup> and SLaM <sup>b</sup>     | SLaM approach: $F_1$ -score 0.85 versus 0.68; WCM approach: $F_1$ -score 0.87 versus 0.72                                            |
| Traditional machine learning |                                                                                                                        |                                                                     |                                                                                                                                      |
| [30]                         | Predict postdischarge suicides                                                                                         | 448,788 VA <sup>c</sup> patients                                    | AUROC <sup>d</sup> : 0.747-0.780                                                                                                     |
| [46]                         | Develop computational phenotypes for patients with opioid-related disorders                                            | 82,577 patients from 10 sites within a regional health care network | Reveal 9 distinct opioid-related cohorts                                                                                             |
| Hybrid                       |                                                                                                                        |                                                                     |                                                                                                                                      |
| [57]                         | Design a data extraction strategy for 21 common physical comorbidities                                                 | 17,500 individuals at SLaM                                          | Precision rates ( $F_1$ -score) above 0.90 for all conditions                                                                        |
| [73]                         | Detect delirium episodes                                                                                               | 1,565,678 clinical notes from 10,516 patients from 9 hospitals      | Micro $F_1$ -score=0.978; macro $F_1$ -score=0.918                                                                                   |
| Deep learning                |                                                                                                                        |                                                                     |                                                                                                                                      |
| [39]                         | Predict mental health crises                                                                                           | 59,750 patients from NHS <sup>e</sup>                               | AUROC: 0.865                                                                                                                         |
| [58]                         | Screen for substance misuse                                                                                            | 54,915 and 1991 patients at RUMC and LUMC                           | AUROC: 0.88 for alcohol misuse; 0.94 for opioid misuse                                                                               |
| Large language models-based  |                                                                                                                        |                                                                     |                                                                                                                                      |
| [65]                         | Evaluate if AI <sup>f</sup> -generated psychiatric discharge summaries match the quality of those written by residents | 20 cases at the Psychiatric University Hospital Zurich              | Humans scored significantly higher (3.78 vs 3.12, $P<.05$ ); Found hallucinations in 40% of AI summaries (37.5% clinically relevant) |
| [64]                         | Develop and evaluate a framework to classify CI stages                                                                 | 1002 & 769 patients at MGB                                          | GPT-4o achieved high accuracy (Weighted Kappa 0.95), outperforming BERT and USE models                                               |

<sup>a</sup>WCM: Weill Cornell Medicine.

<sup>b</sup>SLaM: South London and Maudsley.

<sup>c</sup>VA: Department of Veterans Affairs.

<sup>d</sup>AUROC: area under the receiver operating characteristic curve.

<sup>e</sup>NHS: National Health Service.

<sup>f</sup>AI: artificial intelligence.

## Discussion

### Principal Results

The application of NLP to psychiatric clinical notes presents unique challenges and opportunities, distinct from other medical domains. This stems from the inherently narrative and subjective nature of mental health documentation, where nuanced descriptions of patients' experiences and clinicians' interpretations take precedence over objective measurements. Psychiatric data is often dense, nuanced, and sensitive. Consequently, psychiatric clinical notes demand specialized NLP approaches capable of maintaining context over extensive narratives and extracting meaningful insights from highly subjective content. Parsing these nuanced expressions is crucial for capturing the complexity of a patient's mental state, which often defies straightforward quantification.

Rule-based methods have long been dominant in psychiatric NLP for IE and TC. Their interpretability, customizability, and efficacy with smaller datasets make them suitable for evolving clinical guidelines. However, a significant drawback of such systems lies in their inability to grasp context, particularly when it comes to negations and expressions of uncertainty. As the complexity of mental health data grows, the field has shifted toward hybrid systems that combine the strengths of rule-based methods and ML approaches. Hybrid systems offer enhanced scalability while retaining transparency and adaptability to clinical needs. Despite this, DL has seen limited adoption in psychiatric clinical notes processing due to challenges such as data scarcity and the demand for model interpretability in clinical decision-making.

In recent years, PLMs based on transformers, such as BERT, have become the cornerstone of psychiatric NLP. By learning linguistic representations directly from large corpora, these models capture semantic nuances far better than previous NLP methods. Their value in mental health applications appears to lie particularly in domain adaptation and task-specific fine-tuning. Domain-adapted variants, including MentalBERT and Bio\_ClinicalBERT, extend the utility of general PLMs by continued pretraining on biomedical and clinical corpora, thereby improving the handling of medical terminology, note structure, and psychiatric narrative style. When further fine-tuned for downstream tasks, these models have achieved state-of-the-art performance in tasks such as phenotyping [69], outcome prediction [63], and extracting features [62] from psychiatric clinical notes. However, these advances come with challenges when applied to long longitudinal records, including difficulty handling the exponentially increasing computing requirements in response to the input length.

LLMs offer promising advancements for psychiatric NLP, marking a significant paradigm shift from traditional discriminative tasks to generative and reasoning-based applications. With their flexibility and zero-shot capabilities, LLMs can transform clinical workflows by automating routine tasks, such as generating expert-level discharge summaries [65]. Furthermore, CoT prompting enables the simulation of complex clinical reasoning. Recent studies suggest that LLMs' use in mental health is expanding beyond summarization to broader forms of clinical assistance [179], including automated coding of clinical encounters, real-time clinical decision support, and the generation of synthetic high-fidelity patient profiles [66]. However, in psychiatric contexts, the risk of hallucination is particularly consequential: models may generate plausible but incorrect interpretations of symptoms, suicidality, psychosis, or treatment adherence, thereby introducing errors into clinical documentation or downstream decision-making. More broadly, the high output variability and limited inherent explainability of LLMs remain major barriers to safe clinical deployment [180]. For this reason, future psychiatric LLM systems will likely require not only retrieval-augmented generation [181], but also stronger human oversight, structured validation, and privacy-preserving deployment strategies [182]. In addition, the increasing reliance on proprietary closed-source models raises important ethical concerns in psychiatry, where clinical notes often contain highly sensitive and stigmatizing information. This consideration has strengthened interest in more transparent and locally deployable alternatives. In particular, parameter-efficient techniques such as LoRA make it possible for smaller fine-tuned models to achieve performance comparable to large proprietary models at a fraction of the cost [67].

Our review reveals a clear evolutionary trajectory in psychiatric NLP methods: from rule-based systems to ML, then to hybrid and DL, and most recently to LLM-based approaches. Each paradigm shift has expanded the capabilities of the field while introducing new trade-offs, as summarized in Table 6. Whereas rule-based methods and traditional ML prioritize interpretability and work well with limited labeled data, DL offers superior contextual understanding at the cost of transparency and data requirements. Hybrid methods help bridge this gap by domain knowledge integration, although they often increase system complexity and maintenance burden. LLMs extend this further by enabling generative applications and few-shot learning, but introduce new concerns around factual reliability and clinical safety. The choice of approach should therefore be guided by the specific clinical task, available resources, data characteristics, and the acceptable trade-off between model performance and interpretability.

**Table 6.** Comparative advantages and disadvantages of natural language processing approaches for psychiatric clinical notes.

| NLP <sup>a</sup> approaches | Advantages                                                                                                                                                                                                                       | Disadvantages                                                                                                                                                                                      |
|-----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Rule-based                  | <ul style="list-style-type: none"><li>• High interpretability</li><li>• No training data needed</li><li>• Easy to audit</li><li>• Effective with expert curation</li></ul>                                                       | <ul style="list-style-type: none"><li>• Labor-intensive rule creation</li><li>• Poor scalability</li><li>• Brittle to linguistic variation</li><li>• Cannot capture implicit semantics</li></ul>   |
| Traditional ML <sup>b</sup> | <ul style="list-style-type: none"><li>• Learn patterns from data</li><li>• Rich feature representations</li><li>• Interpretable feature importance</li><li>• Effective on smaller datasets</li></ul>                             | <ul style="list-style-type: none"><li>• Requires feature engineering</li><li>• Performance depends on feature quality</li><li>• Limited long-range context modeling</li></ul>                      |
| Hybrid                      | <ul style="list-style-type: none"><li>• Combine rule interpretability with ML adaptability</li><li>• Balance precision and recall</li><li>• Modular and extensible</li></ul>                                                     | <ul style="list-style-type: none"><li>• Increased system complexity</li><li>• Requires both domain expertise and ML skills</li><li>• Maintenance overhead</li></ul>                                |
| Deep learning               | <ul style="list-style-type: none"><li>• Automatic feature learning</li><li>• Capture contextual semantics</li><li>• Domain-adapted BERT<sup>c</sup> available</li><li>• Strong generalization ability</li></ul>                  | <ul style="list-style-type: none"><li>• “Black-box” nature</li><li>• Higher computational cost</li><li>• Performance suffers with limited data</li><li>• Difficulty with long documents</li></ul>  |
| LLMs <sup>d</sup> -based    | <ul style="list-style-type: none"><li>• Zero and few-shot capability</li><li>• Chain-of-thought reasoning</li><li>• High computational efficiency</li><li>• Versatile across tasks</li><li>• Synthetic data generation</li></ul> | <ul style="list-style-type: none"><li>• Hallucination risk</li><li>• Prompt sensitivity</li><li>• High computational demands</li><li>• Privacy concerns</li><li>• Regulatory uncertainty</li></ul> |

<sup>a</sup>NLP: natural language processing.  
<sup>b</sup>ML: machine learning.  
<sup>c</sup>BERT: Bidirectional Encoder Representations from Transformers.  
<sup>d</sup>LLM: large language model.

The temporal dynamics of mental health conditions further complicate NLP applications in psychiatry. Many psychiatric disorders follow nonlinear progressions, necessitating sophisticated temporal modeling techniques that can track both rapid behavioral changes and long-term symptom trajectories [37,49]. This temporal aspect is vital in psychiatry, where the course of illness can be as informative as the symptoms themselves [183]. Notably, recent NLP research has expanded beyond single-disorder tracking to transdiagnostic approaches. For instance, dynamic temporal network analysis has been successfully applied to model the prodrome of severe mental disorders [51]. By mapping causal pathways among NLP-derived features, researchers identified distinct behavioral communities that consistently precede the onset of full-threshold disorders.

Multimodal integration is another hallmark of psychiatric NLP. Mental health is influenced by a complex interplay of biological, psychological, and social factors, necessitating NLP approaches that can synthesize information from diverse sources, including clinical notes [31,45], genetic studies [184], neuroimaging results [185,186], and even social media interactions [187]. This holistic approach represents a significant departure from traditional, siloed methods of medical data analysis. Integrating NLP-derived clinical features with brain MRI has proven effective in predicting complex conditions like treatment-resistant depression [74]; combining text-based symptom extraction with brain network analysis yields superior predictive performance compared to unimodal methods, validating the additive value of integration.

Public datasets like MIMIC-III are scarce in mental health, meaning most research relies on proprietary institutional data. Research in this domain is distributed across a wide range of institutions, particularly in the US and UK, with key contributors such as major medical centers like SLaM, the VA, and Rush University Medical Center focusing on substance use. However, limited data access is only part of the challenge. Psychiatric clinical notes also reflect important institutional and population-specific differences that may affect model development and generalizability. Clinical notes can vary across institutions in charting style, documentation granularity, terminology, and local diagnostic practice, which may influence the linguistic patterns captured by NLP models. In addition, datasets from individual hospitals or health systems often reflect the demographic, socioeconomic, and clinical characteristics of their local patient populations, introducing potential biases that may reduce the portability of findings across settings. These factors complicate cross-institutional collaboration and create technical barriers to external validation, transfer learning, and domain adaptation. While general health care standards such as OMOP and HL7 FHIR could support improved interoperability, their adoption in psychiatry remains limited, and structural harmonization alone may be insufficient to address variation in narrative documentation. To address these challenges, data-sharing initiatives, privacy-preserving methods such as federated learning, and more systematic cross-site evaluation will be important.

Subjectivity in psychiatric assessments introduces a level of variability that profoundly impacts NLP model



development and evaluation. Unlike other medical fields with objective biomarker-based standards, psychiatric NLP must contend with inter-rater variability among clinicians. This challenge has spurred the development of novel evaluation metrics, such as clinician agreement rates and severity-aware measures, particularly crucial in high-stakes scenarios like suicide risk assessment [55].

## Limitations

Although this scoping review was conducted according to the PRISMA-ScR guidelines and a rigorous search strategy, there were some limitations that are worth noting. First, consistent with scoping review methodology, we did not perform a formal risk-of-bias assessment because the aim of this review was to map the scope and methodological landscape of the field rather than to synthesize effect estimates. Second, the requirement for terms specifically related to “psychiatry” or “psychiatric disorder” in the title or abstract means it is possible that some relevant articles focusing on specific conditions without using the broader umbrella terms were not included in this review. Third, the exclusion of non-English text limits the scope of this review to Anglophone clinical settings, potentially overlooking valuable methodological

developments in other languages. Finally, this review was conducted in an area of research that is constantly growing and developing and therefore only provides a time-stamped representation of the field.

## Conclusions

Psychiatric NLP is shaped by the distinctive characteristics of mental health documentation: subjective narrative language, longitudinal complexity, and heightened privacy and ethical stakes. The field is progressing from interpretable rule-based systems toward hybrid methods and Transformer-based PLMs, with LLMs enabling new generative and reasoning-based workflows. As these tools evolve, they hold the potential to significantly improve patient care, advance our understanding of mental illness, and ultimately alleviate the burden of those living with mental health conditions. The path forward requires collaborative efforts across health care systems, development of adaptive learning models, and careful ethical implementation. In addressing these challenges, psychiatric NLP not only promises to transform mental health care but also to expand the frontiers of clinical NLP more broadly.

## Funding

This work was supported by STI2030-Major Projects (2022ZD0212400, 2021ZD0200404), National Natural Science Foundation of China (82371453), Zhejiang Key Laboratory of Clinical and Basic Research for Psychiatric Diseases (2024ZY01010, 2024E10107), "Pioneer" and "Leading Goose" R&D Program of Zhejiang (2024C03006, 2024C04024, 2026C01013), Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (JYB2025XDXM605), Leading Innovation and Entrepreneurship Team of Zhejiang Province (2023R01005), the Leading Innovation and Entrepreneurship Team of Hangzhou city (TD2024003) and the Construction Fund of Key Medical Disciplines of Hangzhou (2025HZGF10).

## Authors' Contributions

Conceptualization, methodology, investigation, data curation, validation, writing – original draft, and writing – review and editing: SR

Investigation, data curation, validation, writing – original draft: XC

Investigation, data curation: TJ, GD, JX

Supervision and writing – review and editing: HJ, YZ

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Search strategy.

[\[DOCX File \(Microsoft Word File\), 17 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Data extraction template.

[\[DOCX File \(Microsoft Word File\), 17 KB-Multimedia Appendix 2\]](#)

## Multimedia Appendix 3

Definition of terms.

[\[DOCX File \(Microsoft Word File\), 17 KB-Multimedia Appendix 3\]](#)

## Multimedia Appendix 4

Included records.

[\[DOCX File \(Microsoft Word File\), 31 KB-Multimedia Appendix 4\]](#)

## Multimedia Appendix 5

Distribution of different data sources.

[[DOCX File \(Microsoft Word File\), 208 KB-Multimedia Appendix 5](#)]

## Checklist 1

PRISMA-ScR Checklist.

[[DOCX File \(Microsoft Word File\), 32 KB-Checklist 1](#)]

## References

1. Rehm J, Shield KD. Global burden of disease and the impact of mental and addictive disorders. *Curr Psychiatry Rep*. Feb 7, 2019;21(2):10. [doi: [10.1007/s11920-019-0997-0](#)] [Medline: [30729322](#)]
2. Santomauro DF, Mantilla Herrera AM, Shadid J, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet*. Nov 6, 2021;398(10312):1700-1712. [doi: [10.1016/S0140-6736\(21\)02143-7](#)] [Medline: [34634250](#)]
3. Insel TR. Digital phenotyping: technology for a new science of behavior. *JAMA*. Oct 3, 2017;318(13):1215-1216. [doi: [10.1001/jama.2017.11295](#)] [Medline: [28973224](#)]
4. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. *Risk Manag Healthc Policy*. 2011;4:47-55. [doi: [10.2147/RMHP.S12985](#)] [Medline: [22312227](#)]
5. Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a focus on meaningful use. *Ann Intern Med*. Jan 7, 2014;160(1):48-54. [doi: [10.7326/M13-1531](#)] [Medline: [24573664](#)]
6. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? *J Biomed Inform*. Oct 2009;42(5):760-772. [doi: [10.1016/j.jbi.2009.08.007](#)] [Medline: [19683066](#)]
7. Manning C, Schütze H. *Foundations of Statistical Natural Language Processing*. MIT Press; 1999. ISBN: 9780262303798
8. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *J Am Med Inform Assoc*. Oct 1, 2018;25(10):1419-1428. [doi: [10.1093/jamia/ocy068](#)] [Medline: [29893864](#)]
9. Wu S, Roberts K, Datta S, et al. Deep learning in clinical natural language processing: a methodical review. *J Am Med Inform Assoc*. Mar 1, 2020;27(3):457-470. [doi: [10.1093/jamia/ocz200](#)] [Medline: [31794016](#)]
10. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; Jun 2-7, 2019:4171-4186; Minneapolis, MN, USA. 2019. [doi: [10.18653/v1/N19-1423](#)]
11. Bubeck S, Chandrasekaran V, Eldan R, et al. Sparks of artificial general intelligence: early experiments with GPT-4. *arXiv*. Preprint posted online on Apr 13, 2023. [doi: [10.48550/arXiv.2303.12712](#)]
12. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature*. Aug 2023;620(7972):172-180. [doi: [10.1038/s41586-023-06291-2](#)] [Medline: [37438534](#)]
13. Kraljevic Z, Bean D, Shek A, et al. Foresight--generative pretrained transformer (GPT) for modelling of patient timelines using EHRs. *arXiv*. Preprint posted online on Jan 24, 2023. [doi: [10.48550/arXiv.2212.08072](#)]
14. Downs J, Velupillai S, George G, et al. Detection of suicidality in adolescents with autism spectrum disorders: developing a natural language processing approach for use in electronic health records. *AMIA Annu Symp Proc*. Apr 16, 2018;2017:641-649. [Medline: [29854129](#)]
15. Kshatriya BSA, Nunez NA, Resendez MG, et al. Neural language models with distant supervision to identify major depressive disorder from clinical notes. *arXiv*. Preprint posted online on Apr 19, 2021. [doi: [10.48550/arXiv.2104.09644](#)]
16. Tran T, Kavuluru R. Predicting mental conditions based on "history of present illness" in psychiatric notes with deep neural networks. *J Biomed Inform*. Nov 2017;75S:S138-S148. [doi: [10.1016/j.jbi.2017.06.010](#)] [Medline: [28606869](#)]
17. Zhang T, Schoene AM, Ji S, Ananiadou S. Natural language processing applied to mental illness detection: a narrative review. *NPJ Digit Med*. Apr 8, 2022;5(1):46. [doi: [10.1038/s41746-022-00589-7](#)] [Medline: [35396451](#)]
18. Malgaroli M, Hull TD, Zech JM, Althoff T. Natural language processing for mental health interventions: a systematic review and research framework. *Transl Psychiatry*. Oct 6, 2023;13(1):309. [doi: [10.1038/s41398-023-02592-2](#)] [Medline: [37798296](#)]
19. Le Glaz A, Haralambous Y, Kim-Dufoir DH, et al. Machine Learning and Natural Language Processing in Mental Health: Systematic Review. *J Med Internet Res*. May 4, 2021;23(5):e15708. [doi: [10.2196/15708](#)] [Medline: [33944788](#)]

20. Jin Y, Liu J, Li P, et al. The applications of large language models in mental health: scoping review. *J Med Internet Res*. May 5, 2025;27:e69284. [doi: [10.2196/69284](https://doi.org/10.2196/69284)] [Medline: [40324177](https://pubmed.ncbi.nlm.nih.gov/40324177/)]
21. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. Oct 2, 2018;169(7):467-473. [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
22. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. Feb 23, 2007;8(1):19-32. [doi: [10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616)]
23. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. *JBHI Evidence Synthesis*. 2020;18(10):2119-2126. [doi: [10.11124/JBIES-20-00167](https://doi.org/10.11124/JBIES-20-00167)]
24. Natural language processing applied to psychiatric clinical notes: scoping review. *Open Science Framework*. Apr 14, 2026. URL: <https://osf.io/yjec6>
25. Malow BA, Veatch OJ, Niu X, et al. A practical approach to identifying autistic adults within the electronic health record. *Autism Res*. Jan 2023;16(1):52-65. [doi: [10.1002/aur.2849](https://doi.org/10.1002/aur.2849)] [Medline: [36377765](https://pubmed.ncbi.nlm.nih.gov/36377765/)]
26. Schirle L, Jeffery A, Yaqoob A, Sanchez-Roige S, Samuels DC. Two data-driven approaches to identifying the spectrum of problematic opioid use: a pilot study within a chronic pain cohort. *Int J Med Inform*. Dec 2021;156:104621. [doi: [10.1016/j.ijmedinf.2021.104621](https://doi.org/10.1016/j.ijmedinf.2021.104621)] [Medline: [34673309](https://pubmed.ncbi.nlm.nih.gov/34673309/)]
27. Afshar M, Adelaine S, Resnik F, et al. Deployment of real-time natural language processing and deep learning clinical decision support in the electronic health record: pipeline implementation for an opioid misuse screener in hospitalized adults. *JMIR Med Inform*. Apr 20, 2023;11:e44977. [doi: [10.2196/44977](https://doi.org/10.2196/44977)] [Medline: [37079367](https://pubmed.ncbi.nlm.nih.gov/37079367/)]
28. Msosa YJ, Grauslys A, Zhou Y, et al. Trustworthy Data and AI Environments for Clinical Prediction: Application to Crisis-Risk in People With Depression. *IEEE J Biomed Health Inform*. Nov 2023;27(11):5588-5598. [doi: [10.1109/JBHI.2023.3312011](https://doi.org/10.1109/JBHI.2023.3312011)] [Medline: [37669205](https://pubmed.ncbi.nlm.nih.gov/37669205/)]
29. Joyce C, Markossian TW, Nikolaides J, et al. The evaluation of a clinical decision support tool using natural language processing to screen hospitalized adults for unhealthy substance use: protocol for a quasi-experimental design. *JMIR Res Protoc*. Dec 19, 2022;11(12):e42971. [doi: [10.2196/42971](https://doi.org/10.2196/42971)] [Medline: [36534461](https://pubmed.ncbi.nlm.nih.gov/36534461/)]
30. Kessler RC, Bauer MS, Bishop TM, et al. Evaluation of a model to target high-risk psychiatric inpatients for an intensive postdischarge suicide prevention intervention. *JAMA Psychiatry*. Mar 1, 2023;80(3). [doi: [10.1001/jamapsychiatry.2022.4634](https://doi.org/10.1001/jamapsychiatry.2022.4634)] [Medline: [36652267](https://pubmed.ncbi.nlm.nih.gov/36652267/)]
31. Zhao M, Havrilla J, Peng J, et al. Development of a phenotype ontology for autism spectrum disorder by natural language processing on electronic health records. *J Neurodev Disord*. May 23, 2022;14(1):32. [doi: [10.1186/s11689-022-09442-0](https://doi.org/10.1186/s11689-022-09442-0)] [Medline: [35606697](https://pubmed.ncbi.nlm.nih.gov/35606697/)]
32. Bulik CM, Bertoia ML, Lu M, Seeger JD, Spalding WM. Suicidality risk among adults with binge-eating disorder. *Suicide Life Threat Behav*. Oct 2021;51(5):897-906. [doi: [10.1111/sltb.12768](https://doi.org/10.1111/sltb.12768)] [Medline: [34080227](https://pubmed.ncbi.nlm.nih.gov/34080227/)]
33. Chatham AH, Bradley ED, Schirle L, Sanchez-Roige S, Samuels DC, Jeffery AD. Detecting problematic opioid use in the electronic health record: automation of the addiction behaviors checklist in a chronic pain population. *medRxiv*. Preprint posted online on Jun 12, 2023. [doi: [10.1101/2023.06.08.23290894](https://doi.org/10.1101/2023.06.08.23290894)] [Medline: [37398208](https://pubmed.ncbi.nlm.nih.gov/37398208/)]
34. Shah-Mohammadi F, Cui W, Bachi K, Hurd Y, Finkelstein J. Using natural language processing of clinical notes to predict outcomes of opioid treatment program. Presented at: 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) 2022; Jul 11-15, 2022:4415-4420; Glasgow, Scotland, United Kingdom. [doi: [10.1109/EMBC48229.2022.9871960](https://doi.org/10.1109/EMBC48229.2022.9871960)]
35. Zhu VJ, Lenert LA, Barth KS, et al. Automatically identifying opioid use disorder in non-cancer patients on chronic opioid therapy. *Health Informatics J*. 2022;28(2):14604582221107808. [doi: [10.1177/14604582221107808](https://doi.org/10.1177/14604582221107808)] [Medline: [35726687](https://pubmed.ncbi.nlm.nih.gov/35726687/)]
36. Spalding WM, Bertoia ML, Bulik CM, Seeger JD. Treatment characteristics among patients with binge-eating disorder: an electronic health records analysis. *Postgrad Med*. Apr 2023;135(3):254-264. [doi: [10.1080/00325481.2021.2018255](https://doi.org/10.1080/00325481.2021.2018255)] [Medline: [35037815](https://pubmed.ncbi.nlm.nih.gov/35037815/)]
37. Young M, Holmes NE, Kishore K, et al. Natural language processing diagnosed behavioural disturbance phenotypes in the intensive care unit: characteristics, prevalence, trajectory, treatment, and outcomes. *Crit Care*. Nov 4, 2023;27(1):425. [doi: [10.1186/s13054-023-04695-0](https://doi.org/10.1186/s13054-023-04695-0)] [Medline: [37925406](https://pubmed.ncbi.nlm.nih.gov/37925406/)]
38. George A, Johnson D, Carenini G, Eslami A, Ng R, Portales-Casamar E. Applications of aspect-based sentiment analysis on psychiatric clinical notes to study suicide in youth. *AMIA Jt Summits Transl Sci Proc*. 2021;2021(101539486):229-237. [Medline: [34457137](https://pubmed.ncbi.nlm.nih.gov/34457137/)]
39. Garriga R, Buda TS, Guerreiro J, Omaña Iglesias J, Estella Aguerri I, Matić A. Combining clinical notes with structured electronic health records enhances the prediction of mental health crises. *Cell Rep Med*. Nov 21, 2023;4(11):101260. [doi: [10.1016/j.xcrm.2023.101260](https://doi.org/10.1016/j.xcrm.2023.101260)] [Medline: [37913776](https://pubmed.ncbi.nlm.nih.gov/37913776/)]

40. Carson NJ, Yang X, Mullin B, et al. Predicting adolescent suicidal behavior following inpatient discharge using structured and unstructured data. *J Affect Disord.* Apr 1, 2024;350:382-387. [doi: [10.1016/j.jad.2023.12.059](https://doi.org/10.1016/j.jad.2023.12.059)] [Medline: [38158050](https://pubmed.ncbi.nlm.nih.gov/38158050/)]
41. Levis M, Levy J, Dent KR, et al. Leveraging natural language processing to improve electronic health record suicide risk prediction for veterans health administration users. *J Clin Psychiatry.* Jun 19, 2023;84(4):22m14568. [doi: [10.4088/JCP.22m14568](https://doi.org/10.4088/JCP.22m14568)] [Medline: [37341477](https://pubmed.ncbi.nlm.nih.gov/37341477/)]
42. Chen J, Engelhard M, Henao R, et al. Enhancing early autism prediction based on electronic records using clinical narratives. *J Biomed Inform.* Aug 2023;144:104390. [doi: [10.1016/j.jbi.2023.104390](https://doi.org/10.1016/j.jbi.2023.104390)] [Medline: [37182592](https://pubmed.ncbi.nlm.nih.gov/37182592/)]
43. Morrow D, Zamora-Resendiz R, Beckham JC, et al. A case for developing domain-specific vocabularies for extracting suicide factors from healthcare notes. *J Psychiatr Res.* Jul 2022;151:328-338. [doi: [10.1016/j.jpsychires.2022.04.009](https://doi.org/10.1016/j.jpsychires.2022.04.009)] [Medline: [35533516](https://pubmed.ncbi.nlm.nih.gov/35533516/)]
44. Levis M, Levy J, Dufort V, Gobbel GT, Watts BV, Shiner B. Leveraging unstructured electronic medical record notes to derive population-specific suicide risk models. *Psychiatry Res.* Sep 2022;315:114703. [doi: [10.1016/j.psychres.2022.114703](https://doi.org/10.1016/j.psychres.2022.114703)] [Medline: [35841702](https://pubmed.ncbi.nlm.nih.gov/35841702/)]
45. D'Souza EW, MacGregor AJ, Markwald RR, Elkins TA, Zouris JM. Investigating insomnia in United States deployed military forces: a topic modeling approach. *Sleep Health.* Feb 2024;10(1):75-82. [doi: [10.1016/j.sleh.2023.09.014](https://doi.org/10.1016/j.sleh.2023.09.014)] [Medline: [38071173](https://pubmed.ncbi.nlm.nih.gov/38071173/)]
46. Taylor RA, Gilson A, Schulz W, et al. Computational phenotypes for patients with opioid-related disorders presenting to the emergency department. *PLoS One.* Sep 15, 2023;18(9):e0291572. [doi: [10.1371/journal.pone.0291572](https://doi.org/10.1371/journal.pone.0291572)] [Medline: [37713393](https://pubmed.ncbi.nlm.nih.gov/37713393/)]
47. Lee DY, Kim N, Park C, et al. Explainable multimodal prediction of treatment-resistance in patients with depression leveraging brain morphometry and natural language processing. *Psychiatry Res.* Apr 2024;334:115817. [doi: [10.1016/j.psychres.2024.115817](https://doi.org/10.1016/j.psychres.2024.115817)] [Medline: [38430816](https://pubmed.ncbi.nlm.nih.gov/38430816/)]
48. Levis M, Levy J, Dimambro M, et al. Using natural language processing to evaluate temporal patterns in suicide risk variation among high-risk veterans. *Psychiatry Res.* Sep 2024;339:116097. [doi: [10.1016/j.psychres.2024.116097](https://doi.org/10.1016/j.psychres.2024.116097)] [Medline: [39083961](https://pubmed.ncbi.nlm.nih.gov/39083961/)]
49. Oh IY, Schindler SE, Ghoshal N, Lai AM, Payne PRO, Gupta A. Extraction of clinical phenotypes for Alzheimer's disease dementia from clinical notes using natural language processing. *JAMIA Open.* Apr 2023;6(1):o0ad014. [doi: [10.1093/jamiaopen/o0ad014](https://doi.org/10.1093/jamiaopen/o0ad014)] [Medline: [36844369](https://pubmed.ncbi.nlm.nih.gov/36844369/)]
50. Arribas M, Oliver D, Patel R, et al. A transdiagnostic prodrome for severe mental disorders: an electronic health record study. *Mol Psychiatry.* Nov 2024;29(11):3305-3315. [doi: [10.1038/s41380-024-02533-5](https://doi.org/10.1038/s41380-024-02533-5)] [Medline: [38710907](https://pubmed.ncbi.nlm.nih.gov/38710907/)]
51. Arribas M, Barnby JM, Patel R, et al. Longitudinal evolution of the transdiagnostic prodrome to severe mental disorders: a dynamic temporal network analysis informed by natural language processing and electronic health records. *Mol Psychiatry.* Jul 2025;30(7):2931-2942. [doi: [10.1038/s41380-025-02896-3](https://doi.org/10.1038/s41380-025-02896-3)] [Medline: [39843546](https://pubmed.ncbi.nlm.nih.gov/39843546/)]
52. Krakowski K, Oliver D, Arribas M, Stahl D, Fusar-Poli P. Dynamic and transdiagnostic risk calculator based on natural language processing for the prediction of psychosis in secondary mental health care: development and internal-external validation cohort study. *Biol Psychiatry.* Oct 1, 2024;96(7):604-614. [doi: [10.1016/j.biopsych.2024.05.022](https://doi.org/10.1016/j.biopsych.2024.05.022)] [Medline: [38852896](https://pubmed.ncbi.nlm.nih.gov/38852896/)]
53. Hutto A, Zikry TM, Bohac B, et al. Using a natural language processing toolkit to classify electronic health records by psychiatric diagnosis. *Health Informatics J.* 2024;30(4):39466373. [doi: [10.1177/14604582241296411](https://doi.org/10.1177/14604582241296411)] [Medline: [39466373](https://pubmed.ncbi.nlm.nih.gov/39466373/)]
54. Wang T, Codling D, Msosa YJ, et al. VIEWER: an extensible visual analytics framework for enhancing mental healthcare. *J Am Med Inform Assoc.* Jan 1, 2026;33(1):144-158. [doi: [10.1093/jamia/ocaf010](https://doi.org/10.1093/jamia/ocaf010)] [Medline: [39847478](https://pubmed.ncbi.nlm.nih.gov/39847478/)]
55. Wang T, Codling D, Bhugra D, et al. Unraveling ethnic disparities in antipsychotic prescribing among patients with psychosis: a retrospective cohort study based on electronic clinical records. *Schizophr Res.* Oct 2023;260:168-179. [doi: [10.1016/j.schres.2023.08.024](https://doi.org/10.1016/j.schres.2023.08.024)] [Medline: [37669576](https://pubmed.ncbi.nlm.nih.gov/37669576/)]
56. Koleck TA, Tatonetti NP, Bakken S, et al. Identifying symptom information in clinical notes using natural language processing. *Nurs Res.* 2021;70(3):173-183. [doi: [10.1097/NNR.0000000000000488](https://doi.org/10.1097/NNR.0000000000000488)] [Medline: [33196504](https://pubmed.ncbi.nlm.nih.gov/33196504/)]
57. Bendayan R, Kraljevic Z, Shaari S, et al. Mapping multimorbidity in individuals with schizophrenia and bipolar disorders: evidence from the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLAM BRC) case register. *BMJ Open.* Jan 24, 2022;12(1):e054414. [doi: [10.1136/bmjopen-2021-054414](https://doi.org/10.1136/bmjopen-2021-054414)] [Medline: [35074819](https://pubmed.ncbi.nlm.nih.gov/35074819/)]
58. Afshar M, Sharma B, Dligach D, et al. Development and multimodal validation of a substance misuse algorithm for referral to treatment using artificial intelligence (SMART-AI): a retrospective deep learning study. *Lancet Digit Health.* Jun 2022;4(6):e426-e435. [doi: [10.1016/S2589-7500\(22\)00041-3](https://doi.org/10.1016/S2589-7500(22)00041-3)] [Medline: [35623797](https://pubmed.ncbi.nlm.nih.gov/35623797/)]



59. Verter V, E F, Frank D, Georghiou A. Text mining of outpatient narrative notes to predict the risk of psychiatric hospitalization. *Transl Psychiatry*. Feb 20, 2025;15(1):60. [doi: [10.1038/s41398-025-03276-9](https://doi.org/10.1038/s41398-025-03276-9)] [Medline: [39979298](https://pubmed.ncbi.nlm.nih.gov/39979298/)]
60. Ford E, Stone J, Oliver D, et al. Local adaptation and validation of a transdiagnostic risk calculator for first episode psychosis using mental health patient records. *Front Psychiatry*. 2025;16:1584719. [doi: [10.3389/fpsyt.2025.1584719](https://doi.org/10.3389/fpsyt.2025.1584719)] [Medline: [40766924](https://pubmed.ncbi.nlm.nih.gov/40766924/)]
61. Todorovic A, Craig P, Pillinger S, et al. Akrivia Health Database-deep patient characterisation using a secondary mental healthcare dataset in England and Wales: cohort profile. *BMJ Open*. Oct 17, 2024;14(10):e088166. [doi: [10.1136/bmjopen-2024-088166](https://doi.org/10.1136/bmjopen-2024-088166)] [Medline: [39419624](https://pubmed.ncbi.nlm.nih.gov/39419624/)]
62. Kulkarni D, Ghosh A, Girdhari A, et al. Enhancing pre-trained contextual embeddings with triplet loss as an effective fine-tuning method for extracting clinical features from electronic health record derived mental health clinical notes. *Nat Lang Process J*. Mar 2024;6:100045. [doi: [10.1016/j.nlp.2023.100045](https://doi.org/10.1016/j.nlp.2023.100045)]
63. Xie K, Gallagher RS, Shinohara RT, et al. Long-term epilepsy outcome dynamics revealed by natural language processing of clinic notes. *Epilepsia*. Jul 2023;64(7):1900-1909. [doi: [10.1111/epi.17633](https://doi.org/10.1111/epi.17633)] [Medline: [37114472](https://pubmed.ncbi.nlm.nih.gov/37114472/)]
64. Leng Y, He Y, Amini S, et al. A GPT-4o-powered framework for identifying cognitive impairment stages in electronic health records. *NPJ Digit Med*. Jul 3, 2025;8(1):401. [doi: [10.1038/s41746-025-01834-5](https://doi.org/10.1038/s41746-025-01834-5)] [Medline: [40610683](https://pubmed.ncbi.nlm.nih.gov/40610683/)]
65. Schwieger A, Angst K, de Bardeci M, et al. Large language models can support generation of standardized discharge summaries - a retrospective study utilizing ChatGPT-4 and electronic health records. *Int J Med Inform*. Dec 2024;192:105654. [doi: [10.1016/j.ijmedinf.2024.105654](https://doi.org/10.1016/j.ijmedinf.2024.105654)] [Medline: [39437512](https://pubmed.ncbi.nlm.nih.gov/39437512/)]
66. Warner A, LeDue J, Cao Y, Tham J, Murphy TH. Synthetic patient and interview transcript creator: an essential tool for LLMs in mental health. *Front Digit Health*. Sep 11, 2025;7:1625444. [doi: [10.3389/fdgth.2025.1625444](https://doi.org/10.3389/fdgth.2025.1625444)] [Medline: [41019286](https://pubmed.ncbi.nlm.nih.gov/41019286/)]
67. Gireesh H, Shukla L, Shivaprakash P, Mukherjee A, Chand P, Murthy P. Language models for standardising clinical notes and information extraction in addiction psychiatry-an empirical study. *Drug Alcohol Rev*. Jan 2026;45(1):e70059. [doi: [10.1111/dar.70059](https://doi.org/10.1111/dar.70059)] [Medline: [41158037](https://pubmed.ncbi.nlm.nih.gov/41158037/)]
68. Krishnamoorthy Srinivasan SS, Bahadur A, Singh S, et al. Demystifying mental health reports through an LLM-based approach. Presented at: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25); Apr 26 to May 1, 2025; Yokohama Japan. [doi: [10.1145/3706599.3720208](https://doi.org/10.1145/3706599.3720208)]
69. Hua Y, Blackley S, Shinn A, Skinner J, Moran L, Zhou L. Identifying psychosis episodes in psychiatric admission notes via rule-based methods, machine learning, and pre-trained language models. *Res Sq*. Preprint posted online on Mar 21, 2024. [doi: [10.21203/rs.3.rs-4126574/v1](https://doi.org/10.21203/rs.3.rs-4126574/v1)] [Medline: [38562731](https://pubmed.ncbi.nlm.nih.gov/38562731/)]
70. Patra BG, Lepow LA, Kasi Reddy Jagadeesh Kumar P, et al. Extracting social support and social isolation information from clinical psychiatry notes: comparing a rule-based natural language processing system and a large language model. *J Am Med Inform Assoc*. Jan 1, 2025;32(1):218-226. [doi: [10.1093/jamia/ocae260](https://doi.org/10.1093/jamia/ocae260)] [Medline: [39423850](https://pubmed.ncbi.nlm.nih.gov/39423850/)]
71. Cliffe C, Cusick M, Vellupillai S, et al. A multisite comparison using electronic health records and natural language processing to identify the association between suicidality and hospital readmission amongst patients with eating disorders. *Int J Eat Disord*. Aug 2023;56(8):1581-1592. [doi: [10.1002/eat.23980](https://doi.org/10.1002/eat.23980)] [Medline: [37194359](https://pubmed.ncbi.nlm.nih.gov/37194359/)]
72. Castro VM, Rosand J, Giacino JT, McCoy TH, Perlis RH. Case-control study of neuropsychiatric symptoms in electronic health records following COVID-19 hospitalization in 2 academic health systems. *Mol Psychiatry*. Sep 2022;27(9):3898-3903. [doi: [10.1038/s41380-022-01646-z](https://doi.org/10.1038/s41380-022-01646-z)] [Medline: [35705635](https://pubmed.ncbi.nlm.nih.gov/35705635/)]
73. Ge W, Alabsi H, Jain A, et al. Identifying patients with delirium based on unstructured clinical notes: observational study. *JMIR Form Res*. Jun 24, 2022;6(6):e33834. [doi: [10.2196/33834](https://doi.org/10.2196/33834)] [Medline: [35749214](https://pubmed.ncbi.nlm.nih.gov/35749214/)]
74. Pozuelo Moyano B, Orgeta V, von Gunten A, et al. Treatment-resistant late-life depression prevalence and clinical/sociodemographic correlates: an electronic health records study. *J Affect Disord*. Jul 15, 2025;381:77-83. [doi: [10.1016/j.jad.2025.03.169](https://doi.org/10.1016/j.jad.2025.03.169)] [Medline: [40174785](https://pubmed.ncbi.nlm.nih.gov/40174785/)]
75. Annapragada AV, Donaruma-Kwoh MM, Annapragada AV, Starosolski ZA. A natural language processing and deep learning approach to identify child abuse from pediatric electronic medical records. *PLOS ONE*. 2021;16(2):e0247404. [doi: [10.1371/journal.pone.0247404](https://doi.org/10.1371/journal.pone.0247404)] [Medline: [33635890](https://pubmed.ncbi.nlm.nih.gov/33635890/)]
76. Coon H, Shabalin AA, DiBlasi E, et al. Absence of nonfatal suicidal behavior preceding suicide death reveals differences in clinical risks. *Psychiatry Res*. May 2025;347:116391. [doi: [10.1016/j.psychres.2025.116391](https://doi.org/10.1016/j.psychres.2025.116391)] [Medline: [40020535](https://pubmed.ncbi.nlm.nih.gov/40020535/)]
77. Khapre S, Stewart R, Taylor C. An evaluation of symptom domains in the 2 years before pregnancy as predictors of relapse in the perinatal period in women with severe mental illness. *Eur Psychiatry*. Mar 19, 2021;64(1):e26. [doi: [10.1192/j.eurpsy.2021.18](https://doi.org/10.1192/j.eurpsy.2021.18)] [Medline: [33736723](https://pubmed.ncbi.nlm.nih.gov/33736723/)]
78. John D, Montvida O, Chin KL, Khunti K, Paul SK. Antidepressant prescriptions and therapy intensification in men and women newly diagnosed with depression in the UK. *J Psychiatr Res*. Oct 2022;154:167-174. [doi: [10.1016/j.jpsychires.2022.06.054](https://doi.org/10.1016/j.jpsychires.2022.06.054)] [Medline: [35944378](https://pubmed.ncbi.nlm.nih.gov/35944378/)]

79. Mason AJC, Bhavsar V, Botelle R, et al. Applying neural network algorithms to ascertain reported experiences of violence in routine mental healthcare records and distributions of reports by diagnosis. *Front Psychiatry*. 2024;15:1181739. [doi: [10.3389/fpsyt.2024.1181739](https://doi.org/10.3389/fpsyt.2024.1181739)] [Medline: [39319350](https://pubmed.ncbi.nlm.nih.gov/39319350/)]
80. Chatham AH, Bradley ED, Troiani V, et al. Automating the addiction behaviors checklist for problematic opioid use identification. *JAMA Psychiatry*. Jun 1, 2025;82(6):591-598. [doi: [10.1001/jamapsychiatry.2025.0424](https://doi.org/10.1001/jamapsychiatry.2025.0424)] [Medline: [40202749](https://pubmed.ncbi.nlm.nih.gov/40202749/)]
81. Thompson HM, Sharma B, Bhalla S, et al. Bias and fairness assessment of a natural language processing opioid misuse classifier: detection and mitigation of electronic health record data disadvantages across racial subgroups. *J Am Med Inform Assoc*. Oct 12, 2021;28(11):2393-2403. [doi: [10.1093/jamia/ocab148](https://doi.org/10.1093/jamia/ocab148)] [Medline: [34383925](https://pubmed.ncbi.nlm.nih.gov/34383925/)]
82. Werbeloff N, Hilge Thygesen J, Hayes JF, Viding EM, Johnson S, Osborn DPJ. Childhood sexual abuse in patients with severe mental illness: demographic, clinical and functional correlates. *Acta Psychiatr Scand*. Jun 2021;143(6):495-502. [doi: [10.1111/acps.13302](https://doi.org/10.1111/acps.13302)] [Medline: [33779997](https://pubmed.ncbi.nlm.nih.gov/33779997/)]
83. Han S, Zhang RF, Shi L, et al. Classifying social determinants of health from unstructured electronic health records using deep learning-based natural language processing. *J Biomed Inform*. Mar 2022;127:103984. [doi: [10.1016/j.jbi.2021.103984](https://doi.org/10.1016/j.jbi.2021.103984)] [Medline: [35007754](https://pubmed.ncbi.nlm.nih.gov/35007754/)]
84. Carrell DS, Cronkite DJ, Shea M, et al. Clinical documentation of patient-reported medical cannabis use in primary care: toward scalable extraction using natural language processing methods. *Subst Abus*. 2022;43(1):917-924. [doi: [10.1080/08897077.2021.1986767](https://doi.org/10.1080/08897077.2021.1986767)] [Medline: [35254218](https://pubmed.ncbi.nlm.nih.gov/35254218/)]
85. Maserejian N, Krzywy H, Eaton S, Galvin JE. Cognitive measures lacking in EHR prior to dementia or Alzheimer's disease diagnosis. *Alzheimer's & Dementia*. Jul 2021;17(7):1231-1243. [doi: [10.1002/alz.12280](https://doi.org/10.1002/alz.12280)]
86. Barbour K, Tian N, Yozawitz EG, et al. Creating rare epilepsy cohorts using keyword search in electronic health records. *Epilepsia*. Oct 2023;64(10):2738-2749. [doi: [10.1111/epi.17725](https://doi.org/10.1111/epi.17725)] [Medline: [37498137](https://pubmed.ncbi.nlm.nih.gov/37498137/)]
87. Noori A, Magdamo C, Liu X, et al. Development and evaluation of a natural language processing annotation tool to facilitate phenotyping of cognitive status in electronic health records: diagnostic study. *J Med Internet Res*. Aug 30, 2022;24(8):e40384. [doi: [10.2196/40384](https://doi.org/10.2196/40384)] [Medline: [36040790](https://pubmed.ncbi.nlm.nih.gov/36040790/)]
88. Wang B, Miller-Fleming TW, Yu D, et al. Development and validation of electronic health record-based ascertainment of obsessive-compulsive disorder cases and controls. *Psychiatry and Clinical Psychology*. [doi: [10.1101/2025.08.05.25332874](https://doi.org/10.1101/2025.08.05.25332874)]
89. Colbert SMC, Lepow L, Fennessy B, et al. Distinguishing clinical and genetic risk factors for suicidal ideation and behavior in a diverse hospital population. *Transl Psychiatry*. Feb 20, 2025;15(1):63. [doi: [10.1038/s41398-025-03287-6](https://doi.org/10.1038/s41398-025-03287-6)] [Medline: [39979244](https://pubmed.ncbi.nlm.nih.gov/39979244/)]
90. Hart KL, Pellegrini AM, Forester BP, et al. Distribution of agitation and related symptoms among hospitalized patients using a scalable natural language processing method. *Gen Hosp Psychiatry*. 2021;68:46-51. [doi: [10.1016/j.genhosppsych.2020.11.003](https://doi.org/10.1016/j.genhosppsych.2020.11.003)] [Medline: [33310013](https://pubmed.ncbi.nlm.nih.gov/33310013/)]
91. Abramsky S, St Rose S, Heng YW, et al. Examining differences in clinical and demographic characteristics of patients with post-traumatic stress disorder across adult treatment subgroups based on the NeuroBlu database: a non-interventional, retrospective cohort study. *BMJ Open*. Nov 9, 2025;15(11):e099711. [doi: [10.1136/bmjopen-2025-099711](https://doi.org/10.1136/bmjopen-2025-099711)] [Medline: [41213678](https://pubmed.ncbi.nlm.nih.gov/41213678/)]
92. Lin Y, Sharma B, Thompson HM, et al. External validation of a machine learning classifier to identify unhealthy alcohol use in hospitalized patients. *Addiction*. Apr 2022;117(4):925-933. [doi: [10.1111/add.15730](https://doi.org/10.1111/add.15730)] [Medline: [34729829](https://pubmed.ncbi.nlm.nih.gov/34729829/)]
93. Afshar M, Sharma B, Bhalla S, et al. External validation of an opioid misuse machine learning classifier in hospitalized adult patients. *Addict Sci Clin Pract*. Mar 17, 2021;16(1):19. [doi: [10.1186/s13722-021-00229-7](https://doi.org/10.1186/s13722-021-00229-7)] [Medline: [33731210](https://pubmed.ncbi.nlm.nih.gov/33731210/)]
94. Wellesley Wesley E, Patel I, Kadra-Scalzo G, et al. Gender disparities in clozapine prescription in a cohort of treatment-resistant schizophrenia in the South London and Maudsley case register. *Schizophr Res*. Jun 2021;232:68-76. [doi: [10.1016/j.schres.2021.05.006](https://doi.org/10.1016/j.schres.2021.05.006)] [Medline: [34022618](https://pubmed.ncbi.nlm.nih.gov/34022618/)]
95. Zolnoori M, Barrón Y, Song J, et al. HomeADScreen: developing Alzheimer's disease and related dementia risk identification model in home healthcare. *Int J Med Inform*. Sep 2023;177:105146. [doi: [10.1016/j.ijmedinf.2023.105146](https://doi.org/10.1016/j.ijmedinf.2023.105146)] [Medline: [37454558](https://pubmed.ncbi.nlm.nih.gov/37454558/)]
96. St Sauver J, Fu S, Sohn S, et al. Identification of delirium from real-world electronic health record clinical notes. *J Clin Transl Sci*. 2023;7(1):e187. [doi: [10.1017/cts.2023.610](https://doi.org/10.1017/cts.2023.610)] [Medline: [37745932](https://pubmed.ncbi.nlm.nih.gov/37745932/)]
97. Stemerman R, Arguello J, Brice J, Krishnamurthy A, Houston M, Kitzmiller R. Identification of social determinants of health using multi-label classification of electronic health record clinical notes. *JAMIA Open*. Jul 2021;4(3):ooaa069. [doi: [10.1093/jamiaopen/ooaa069](https://doi.org/10.1093/jamiaopen/ooaa069)] [Medline: [34514351](https://pubmed.ncbi.nlm.nih.gov/34514351/)]
98. Brown AM, White DG, Adams NB. Identifying co-occurring disorders among patients with an opioid-involved hospital encounter using National Hospital Care Survey data. *Vital Health Stat*. 2022(199):1-20. [doi: [10.15620/cdc.119155](https://doi.org/10.15620/cdc.119155)]

99. Ariño H, Bae SK, Chaturvedi J, Wang T, Roberts A. Identifying encephalopathy in patients admitted to an intensive care unit: going beyond structured information using natural language processing. *Front Digit Health*. 2023;5:1085602. [doi: [10.3389/fdgh.2023.1085602](https://doi.org/10.3389/fdgh.2023.1085602)] [Medline: [36755566](https://pubmed.ncbi.nlm.nih.gov/36755566/)]
100. Datar S, Lindemann EA, Silverman G, et al. Identifying mentions of life stressors in clinical notes. Presented at: 2021 IEEE 9th International Conference on Healthcare Informatics (ICHI; Victoria, BC, Canada. [doi: [10.1109/ICHI52183.2021.00033](https://doi.org/10.1109/ICHI52183.2021.00033)]
101. Xie F, Ling Grant DS, Chang J, Amundsen BI, Hechter RC. Identifying suicidal ideation and attempt from clinical notes within a large integrated health care system. *Perm J*. Apr 5, 2022;26(1):85-93. [doi: [10.7812/TPP/21.102](https://doi.org/10.7812/TPP/21.102)] [Medline: [35609162](https://pubmed.ncbi.nlm.nih.gov/35609162/)]
102. Workman TE, Goulet JL, Brandt CA, et al. Identifying suicide documentation in clinical notes through zero-shot learning. *Health Sci Rep*. Sep 2023;6(9):e1526. [doi: [10.1002/hsr2.1526](https://doi.org/10.1002/hsr2.1526)] [Medline: [37706016](https://pubmed.ncbi.nlm.nih.gov/37706016/)]
103. Shiner B, Levis M, Dufort VM, et al. Improvements to PTSD quality metrics with natural language processing. *J Eval Clin Pract*. Aug 2022;28(4):520-530. [doi: [10.1111/jep.13587](https://doi.org/10.1111/jep.13587)] [Medline: [34028937](https://pubmed.ncbi.nlm.nih.gov/34028937/)]
104. Sheu Y han, Magdamo C, Miller M, Smoller JW, Blacker D. Initial antidepressant choice by non-psychiatrists: learning from large-scale electronic health records. *Gen Hosp Psychiatry*. Mar 2023;81:22-31. [doi: [10.1016/j.genhosppsych.2022.12.004](https://doi.org/10.1016/j.genhosppsych.2022.12.004)]
105. Hanson RF, Zhu V, Are F, et al. Initial development of tools to identify child abuse and neglect in pediatric primary care. *BMC Med Inform Decis Mak*. Nov 17, 2023;23(1):266. [doi: [10.1186/s12911-023-02361-7](https://doi.org/10.1186/s12911-023-02361-7)] [Medline: [37978498](https://pubmed.ncbi.nlm.nih.gov/37978498/)]
106. Sedgwick R, Bittar A, Kalsi H, Barack T, Downs J, Dutta R. Investigating online activity in UK adolescent mental health patients: a feasibility study using a natural language processing approach for electronic health records. *BMJ Open*. May 25, 2023;13(5):e061640. [doi: [10.1136/bmjopen-2022-061640](https://doi.org/10.1136/bmjopen-2022-061640)] [Medline: [37230520](https://pubmed.ncbi.nlm.nih.gov/37230520/)]
107. Burnett A, Chen N, Zeritis S, et al. Machine learning algorithms to classify self-harm behaviours in New South Wales ambulance electronic medical records: a retrospective study. *Int J Med Inform*. May 2022;161:104734. [doi: [10.1016/j.ijmedinf.2022.104734](https://doi.org/10.1016/j.ijmedinf.2022.104734)] [Medline: [35287099](https://pubmed.ncbi.nlm.nih.gov/35287099/)]
108. Hart KL, Perlis RH, McCoy TH. Mapping of transdiagnostic neuropsychiatric phenotypes across patients in two general hospitals. *J Acad Consult Liaison Psychiatry*. 2021;62(4):430-439. [doi: [10.1016/j.jaclp.2021.01.002](https://doi.org/10.1016/j.jaclp.2021.01.002)] [Medline: [34210402](https://pubmed.ncbi.nlm.nih.gov/34210402/)]
109. Seker A, Bullock E, Chandler S, et al. Mood instability as a transdiagnostic predictor of cannabis use in attention-deficit/hyperactivity disorder and depression: a natural language processing analysis of electronic health records from 13,025 adolescents. *Eur Psychiatry*. Aug 22, 2025;68(1):e139. [doi: [10.1192/j.eurpsy.2025.10095](https://doi.org/10.1192/j.eurpsy.2025.10095)] [Medline: [40843511](https://pubmed.ncbi.nlm.nih.gov/40843511/)]
110. Al-Harrasi AM, Iqbal E, Tsamakis K, et al. Motor signs in Alzheimer's disease and vascular dementia: detection through natural language processing, co-morbid features and relationship to adverse outcomes. *Exp Gerontol*. Apr 2021;146:111223. [doi: [10.1016/j.exger.2020.111223](https://doi.org/10.1016/j.exger.2020.111223)] [Medline: [33450346](https://pubmed.ncbi.nlm.nih.gov/33450346/)]
111. Tsui FR, Shi L, Ruiz V, et al. Natural language processing and machine learning of electronic health records for prediction of first-time suicide attempts. *JAMIA Open*. Jan 2021;4(1):ooab011. [doi: [10.1093/jamiaopen/ooab011](https://doi.org/10.1093/jamiaopen/ooab011)] [Medline: [33758800](https://pubmed.ncbi.nlm.nih.gov/33758800/)]
112. Ridgway JP, Uvin A, Schmitt J, et al. Natural language processing of clinical notes to identify mental illness and substance use among people living with HIV: retrospective cohort study. *JMIR Med Inform*. Mar 10, 2021;9(3):e23456. [doi: [10.2196/23456](https://doi.org/10.2196/23456)] [Medline: [33688848](https://pubmed.ncbi.nlm.nih.gov/33688848/)]
113. Wang L, Foer D, MacPhaul E, Lo YC, Bates DW, Zhou L. PASCLeX: A comprehensive post-acute sequelae of COVID-19 (PASC) symptom lexicon derived from electronic health record clinical notes. *J Biomed Inform*. Jan 2022;125:103951. [doi: [10.1016/j.jbi.2021.103951](https://doi.org/10.1016/j.jbi.2021.103951)] [Medline: [34785382](https://pubmed.ncbi.nlm.nih.gov/34785382/)]
114. Cusick M, Velupillai S, Downs J, et al. Portability of natural language processing methods to detect suicidality from clinical text in US and UK electronic health records. *J Affect Disord Rep*. Dec 2022;10:100430. [doi: [10.1016/j.jadr.2022.100430](https://doi.org/10.1016/j.jadr.2022.100430)] [Medline: [36644339](https://pubmed.ncbi.nlm.nih.gov/36644339/)]
115. Vaci N, Koychev I, Kim CH, et al. Real-world effectiveness, its predictors and onset of action of cholinesterase inhibitors and memantine in dementia: retrospective health record study. *Br J Psychiatry*. May 2021;218(5):261-267. [doi: [10.1192/bjp.2020.136](https://doi.org/10.1192/bjp.2020.136)] [Medline: [32713359](https://pubmed.ncbi.nlm.nih.gov/32713359/)]
116. Meerwijk EL, Tamang SR, Finlay AK, Ilgen MA, Reeves RM, Harris AHS. Suicide theory-guided natural language processing of clinical progress notes to improve prediction of veteran suicide risk: protocol for a mixed-method study. *BMJ Open*. Aug 24, 2022;12(8):e065088. [doi: [10.1136/bmjopen-2022-065088](https://doi.org/10.1136/bmjopen-2022-065088)] [Medline: [36002210](https://pubmed.ncbi.nlm.nih.gov/36002210/)]
117. Chilman N, Song X, Roberts A, et al. Text mining occupations from the mental health electronic health record: a natural language processing approach using records from the Clinical Record Interactive Search (CRIS) platform in south London, UK. *BMJ Open*. Mar 25, 2021;11(3):e042274. [doi: [10.1136/bmjopen-2020-042274](https://doi.org/10.1136/bmjopen-2020-042274)] [Medline: [33766838](https://pubmed.ncbi.nlm.nih.gov/33766838/)]
118. Costa T, Menzat B, Engelthaler T, et al. The burden associated with, and management of, difficult-to-treat depression in patients under specialist psychiatric care in the United Kingdom. *J Psychopharmacol*. May 2022;36(5):545-556. [doi: [10.1177/02698811221090628](https://doi.org/10.1177/02698811221090628)] [Medline: [35506640](https://pubmed.ncbi.nlm.nih.gov/35506640/)]

119. Weiner SG, Lo YC, Carroll AD, et al. The incidence and disparities in use of stigmatizing language in clinical notes for patients with substance use disorder. *J Addict Med.* 2023;17(4):424-430. [doi: [10.1097/ADM.0000000000001145](https://doi.org/10.1097/ADM.0000000000001145)] [Medline: [37579100](https://pubmed.ncbi.nlm.nih.gov/37579100/)]
120. Panaite V, Devendorf AR, Finch D, Bouayad L, Luther SL, Schultz SK. The value of extracting clinician-recorded affect for advancing clinical research on depression: proof-of-concept study applying natural language processing to electronic health records. *JMIR Form Res.* May 12, 2022;6(5):e34436. [doi: [10.2196/34436](https://doi.org/10.2196/34436)] [Medline: [35551066](https://pubmed.ncbi.nlm.nih.gov/35551066/)]
121. Bittar A, Velupillai S, Roberts A, Dutta R. Using general-purpose sentiment lexicons for suicide risk assessment in electronic health records: corpus-based analysis. *JMIR Med Inform.* Apr 13, 2021;9(4):e22397. [doi: [10.2196/22397](https://doi.org/10.2196/22397)] [Medline: [33847595](https://pubmed.ncbi.nlm.nih.gov/33847595/)]
122. Cliffe C, Seyedsalehi A, Vardavoulia K, et al. Using natural language processing to extract self-harm and suicidality data from a clinical sample of patients with eating disorders: a retrospective cohort study. *BMJ Open.* Dec 31, 2021;11(12):e053808. [doi: [10.1136/bmjopen-2021-053808](https://doi.org/10.1136/bmjopen-2021-053808)] [Medline: [34972768](https://pubmed.ncbi.nlm.nih.gov/34972768/)]
123. Singleton J, Li C, Akpunonu PD, Abner EL, Kucharska-Newton AM. Using natural language processing to identify opioid use disorder in electronic health record data. *Int J Med Inform.* Feb 2023;170:104963. [doi: [10.1016/j.ijmedinf.2022.104963](https://doi.org/10.1016/j.ijmedinf.2022.104963)] [Medline: [36521420](https://pubmed.ncbi.nlm.nih.gov/36521420/)]
124. Cusick M, Adekanattu P, Campion TR Jr, et al. Using weak supervision and deep learning to classify clinical notes for identification of current suicidal ideation. *J Psychiatr Res.* Apr 2021;136:95-102. [doi: [10.1016/j.jpsychires.2021.01.052](https://doi.org/10.1016/j.jpsychires.2021.01.052)] [Medline: [33581461](https://pubmed.ncbi.nlm.nih.gov/33581461/)]
125. Li Z, Kormilitzin A, Fernandes M, et al. Validation of UK Biobank data for mental health outcomes: a pilot study using secondary care electronic health records. *Int J Med Inform.* Apr 2022;160:104704. [doi: [10.1016/j.ijmedinf.2022.104704](https://doi.org/10.1016/j.ijmedinf.2022.104704)] [Medline: [35168089](https://pubmed.ncbi.nlm.nih.gov/35168089/)]
126. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg.* 2010;8(5):336-341. [doi: [10.1016/j.ijsu.2010.02.007](https://doi.org/10.1016/j.ijsu.2010.02.007)] [Medline: [20171303](https://pubmed.ncbi.nlm.nih.gov/20171303/)]
127. Riehmann P, Hanfler M, Froehlich B. Interactive sankey diagrams. Presented at: Proceedings of the IEEE Symposium on Information Visualization (InfoVis 2005); Oct 23-25, 2005; Minneapolis, MN. [doi: [10.1109/INFOVIS.2005.21](https://doi.org/10.1109/INFOVIS.2005.21)]
128. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data.* May 24, 2016;3:160035. [doi: [10.1038/sdata.2016.35](https://doi.org/10.1038/sdata.2016.35)] [Medline: [27219127](https://pubmed.ncbi.nlm.nih.gov/27219127/)]
129. Perera G, Broadbent M, Callard F, et al. Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register: current status and recent enhancement of an electronic mental health record-derived data resource. *BMJ Open.* Mar 1, 2016;6(3):e008721. [doi: [10.1136/bmjopen-2015-008721](https://doi.org/10.1136/bmjopen-2015-008721)] [Medline: [26932138](https://pubmed.ncbi.nlm.nih.gov/26932138/)]
130. Hoffmire C, Stephens B, Morley S, Thompson C, Kemp J, Bossarte RM. VA Suicide Prevention Applications Network: a national health care system-based suicide event tracking system. *Public Health Rep.* Nov 2016;131(6):816-821. [doi: [10.1177/0033354916670133](https://doi.org/10.1177/0033354916670133)] [Medline: [28123228](https://pubmed.ncbi.nlm.nih.gov/28123228/)]
131. Thompson HM, Faig W, VanKim NA, Sharma B, Afshar M, Karnik NS. Differences in length of stay and discharge destination among patients with substance use disorders: the effect of Substance Use Intervention Team (SUIT) consultation service. *PLoS One.* Oct 9, 2020;15(10):e0239761. [doi: [10.1371/journal.pone.0239761](https://doi.org/10.1371/journal.pone.0239761)] [Medline: [33035229](https://pubmed.ncbi.nlm.nih.gov/33035229/)]
132. Data standardization. OHDSI. URL: <https://www.ohdsi.org/data-standardization/> [Accessed 2024-05-28]
133. BenderD, Sartipik. HL7 FHIR: an agile and restful approach to healthcare information exchange. Presented at: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2013); Jun 20-22, 2013; Porto, Portugal Piscataway, NJ. [doi: [10.1109/CBMS.2013.6627810](https://doi.org/10.1109/CBMS.2013.6627810)]
134. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res.* Jan 1, 2004;32:D267-70. [doi: [10.1093/nar/gkh061](https://doi.org/10.1093/nar/gkh061)] [Medline: [14681409](https://pubmed.ncbi.nlm.nih.gov/14681409/)]
135. Overview of SNOMED CT. National Library of Medicine. URL: [https://www.nlm.nih.gov/healthit/snomedct/snomed\\_overview.html](https://www.nlm.nih.gov/healthit/snomedct/snomed_overview.html) [Accessed 2024-03-12]
136. ICD-10 Version:2019. URL: <https://icd.who.int/browse10/2019/en> [Accessed 2024-03-12]
137. RxNorm overview. National Library of Medicine. URL: <https://www.nlm.nih.gov/research/umls/rxnorm/overview.html> [Accessed 2024-03-12]
138. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. American Psychiatric Association; 2013. [doi: [10.1176/appi.books.9780890425596](https://doi.org/10.1176/appi.books.9780890425596)] ISBN: 9780890425558
139. Wu SM, Compton P, Bolus R, et al. The addiction behaviors checklist: validation of a new clinician-based measure of inappropriate opioid use in chronic pain. *J Pain Symptom Manage.* Oct 2006;32(4):342-351. [doi: [10.1016/j.jpainsymman.2006.05.010](https://doi.org/10.1016/j.jpainsymman.2006.05.010)] [Medline: [17000351](https://pubmed.ncbi.nlm.nih.gov/17000351/)]
140. Clancey WJ. The epistemology of a rule-based expert system — a framework for explanation. *Artif Intell.* May 1983;20(3):215-251. [doi: [10.1016/0004-3702\(83\)90008-5](https://doi.org/10.1016/0004-3702(83)90008-5)]
141. Cowie J, Lehnert W. Information extraction. *Commun ACM.* Jan 1, 1996;39(1):80-91. [doi: [10.1145/234173.234209](https://doi.org/10.1145/234173.234209)]



142. Rindflesch TC, Tanabe L, Weinstein JN, Hunter L. EDGAR: extraction of drugs, genes and relations from the biomedical literature. *Pac Symp Biocomput.* 2000;517-528. [doi: [10.1142/9789814447331\\_0049](https://doi.org/10.1142/9789814447331_0049)] [Medline: [10902199](https://pubmed.ncbi.nlm.nih.gov/10902199/)]
143. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform.* Oct 2001;34(5):301-310. [doi: [10.1006/jbin.2001.1029](https://doi.org/10.1006/jbin.2001.1029)] [Medline: [12123149](https://pubmed.ncbi.nlm.nih.gov/12123149/)]
144. Fu S, Chen D, He H, et al. Clinical concept extraction: a methodology review. *J Biomed Inform.* Sep 2020;109:103526. [doi: [10.1016/j.jbi.2020.103526](https://doi.org/10.1016/j.jbi.2020.103526)] [Medline: [32768446](https://pubmed.ncbi.nlm.nih.gov/32768446/)]
145. Friedman C, Hripcsak G, DuMouchel W, Johnson SB, Clayton PD. Natural language processing in an operational clinical information system. *Nat Lang Eng.* Mar 1995;1(1):83-108. [doi: [10.1017/S1351324900000061](https://doi.org/10.1017/S1351324900000061)]
146. Aronson AR. Effective mapping of biomedical text to the UMLS metathesaurus: the metamap program. Presented at: Proceedings of the AMIA Symposium American Medical Informatics Association; Oct 30 to Nov 3, 2001:17; San Diego, CA, USA. 2001.URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC2243666/> [Accessed 2026-06-30]
147. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc.* 2010;17(5):507-513. [doi: [10.1136/jamia.2009.001560](https://doi.org/10.1136/jamia.2009.001560)] [Medline: [20819853](https://pubmed.ncbi.nlm.nih.gov/20819853/)]
148. Soysal E, Wang J, Jiang M, et al. CLAMP - a toolkit for efficiently building customized clinical natural language processing pipelines. *J Am Med Inform Assoc.* Mar 1, 2018;25(3):331-336. [doi: [10.1093/jamia/ocx132](https://doi.org/10.1093/jamia/ocx132)] [Medline: [29186491](https://pubmed.ncbi.nlm.nih.gov/29186491/)]
149. Nlpie/biomedicus. GitHub. URL: <https://github.com/nlpie/biomedicus> [Accessed 2024-03-12]
150. Reátegui R, Ratté S. Comparison of MetaMap and cTAKES for entity extraction in clinical notes. *BMC Med Inform Decis Mak.* Sep 14, 2018;18(Suppl 3):74. [doi: [10.1186/s12911-018-0654-2](https://doi.org/10.1186/s12911-018-0654-2)] [Medline: [30255810](https://pubmed.ncbi.nlm.nih.gov/30255810/)]
151. Sebastiani F. Machine learning in automated text categorization. *ACM Comput Surv.* Mar 2002;34(1):1-47. [doi: [10.1145/505282.505283](https://doi.org/10.1145/505282.505283)]
152. Pereg O, Korat D, Wasserblat M, Mamou J, Dagan I. ABSApp: a portable weakly-supervised aspect-based sentiment extraction system. *arXiv.* Preprint posted online on Sep 12, 2019. [doi: [10.48550/arXiv.1909.05608](https://doi.org/10.48550/arXiv.1909.05608)]
153. Shickel B, Siegel S, Heesacker M, Benton S, Rashidi P. Automatic detection and classification of cognitive distortions in mental health text. Presented at: 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE); Oct 26-28, 2020; Cincinnati, OH. [doi: [10.1109/BIBE50027.2020.00052](https://doi.org/10.1109/BIBE50027.2020.00052)]
154. Blei DM. Probabilistic topic models. *Commun ACM.* Apr 1, 2012;55(4):77-84. [doi: [10.1145/2133806.2133826](https://doi.org/10.1145/2133806.2133826)]
155. Topaz M, Murga L, Bar-Bachar O, McDonald M, Bowles K. NimbleMiner: an open-source nursing-sensitive natural language processing system based on word embedding. *Comput Inform Nurs.* Nov 2019;37(11):583-590. [doi: [10.1097/CIN.0000000000000557](https://doi.org/10.1097/CIN.0000000000000557)] [Medline: [31478922](https://pubmed.ncbi.nlm.nih.gov/31478922/)]
156. Kraljevic Z, Searle T, Shek A, et al. Multi-domain clinical natural language processing with MedCAT: the medical concept annotation toolkit. *Artif Intell Med.* [doi: [10.1016/j.artmed.2021.102083](https://doi.org/10.1016/j.artmed.2021.102083)] [Medline: [34127232](https://pubmed.ncbi.nlm.nih.gov/34127232/)]
157. Cunningham H, Tablan V, Roberts A, Bontcheva K. Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. *PLoS Comput Biol.* 2013;9(2):e1002854. [doi: [10.1371/journal.pcbi.1002854](https://doi.org/10.1371/journal.pcbi.1002854)] [Medline: [23408875](https://pubmed.ncbi.nlm.nih.gov/23408875/)]
158. Jackson MSc RG, Ball M, Patel R, Hayes RD, Dobson RJB, Stewart R. TextHunter--a user friendly tool for extracting generic concepts from free text in clinical research. *AMIA Annu Symp Proc.* Nov 14, 2014;2014:729-738. [Medline: [25954379](https://pubmed.ncbi.nlm.nih.gov/25954379/)]
159. Jackson RG, Patel R, Jayatilleke N, et al. Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. *BMJ Open.* Jan 17, 2017;7(1):e012012. [doi: [10.1136/bmjopen-2016-012012](https://doi.org/10.1136/bmjopen-2016-012012)] [Medline: [28096249](https://pubmed.ncbi.nlm.nih.gov/28096249/)]
160. Pfaff ER, Crosskey M, Morton K, Krishnamurthy A. Clinical Annotation Research Kit (CLARK): computable phenotyping using machine learning. *JMIR Med Inform.* Jan 24, 2020;8(1):e16042. [doi: [10.2196/16042](https://doi.org/10.2196/16042)] [Medline: [32012059](https://pubmed.ncbi.nlm.nih.gov/32012059/)]
161. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* May 28, 2015;521(7553):436-444. [doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539)] [Medline: [26017442](https://pubmed.ncbi.nlm.nih.gov/26017442/)]
162. Ghannay S, Favre B, Estève Y, Camelin N. Word embedding evaluation and combination. Presented at: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16); May 23-28, 2016:300-305; Portorož, Slovenia. [doi: [10.63317/5monnmod7n4c](https://doi.org/10.63317/5monnmod7n4c)]
163. Lan Z, Chen M, Goodman S, Gimpel K, Sharma P, Soricut R. ALBERT: a lite BERT for self-supervised learning of language representations. *arXiv.* Preprint posted online on Feb 9, 2020. [doi: [10.48550/arXiv.1909.11942](https://doi.org/10.48550/arXiv.1909.11942)]
164. Neumann M, King D, Beltagy I, Ammar W. ScispaCy: fast and robust models for biomedical natural language processing. *arXiv.* Preprint posted online on Oct 9, 2019. [doi: [10.18653/v1/W19-5034](https://doi.org/10.18653/v1/W19-5034)]
165. Kormilitzin A, Vaci N, Liu Q, Nevado-Holgado A. Med7: A transferable clinical natural language processing model for electronic health records. *Artif Intell Med.* Aug 2021;118:102086. [doi: [10.1016/j.artmed.2021.102086](https://doi.org/10.1016/j.artmed.2021.102086)] [Medline: [34412834](https://pubmed.ncbi.nlm.nih.gov/34412834/)]

166. Lee J, Yoon W, Kim S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. Feb 15, 2020;36(4):1234-1240. [doi: [10.1093/bioinformatics/btz682](https://doi.org/10.1093/bioinformatics/btz682)] [Medline: [31501885](https://pubmed.ncbi.nlm.nih.gov/31501885/)]
167. Ji S, Zhang T, Ansari L, Fu J, Tiwari P, Cambria E. MentalBERT: publicly available pretrained language models for mental healthcare. *arXiv*. Preprint posted online on Oct 29, 2021. [doi: [10.48550/arXiv.2110.15621](https://doi.org/10.48550/arXiv.2110.15621)]
168. Xie K, Gallagher RS, Conrad EC, et al. Extracting seizure frequency from epilepsy clinic notes: a machine reading approach to natural language processing. *J Am Med Inform Assoc*. Apr 13, 2022;29(5):873-881. [doi: [10.1093/jamia/ocac018](https://doi.org/10.1093/jamia/ocac018)] [Medline: [35190834](https://pubmed.ncbi.nlm.nih.gov/35190834/)]
169. Alsentzer E, Murphy JR, Boag W, et al. Publicly available clinical BERT embeddings. *arXiv*. Preprint posted online on Jun 20, 2019. [doi: [10.48550/arXiv.1904.03323](https://doi.org/10.48550/arXiv.1904.03323)]
170. Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *J Mach Learn Res*. Jan 1, 2020;21(1):5485-5551. URL: <https://www.jmlr.org/papers/v21/20-074.html> [Accessed 2026-06-30]
171. Kim N, Patel R, Poliak A, et al. Probing what different NLP tasks teach machines about function word comprehension. *arXiv*. Preprint posted online on Aug 7, 2019. [doi: [10.48550/arXiv.1904.11544](https://doi.org/10.48550/arXiv.1904.11544)]
172. Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models. Presented at: *Advances in Neural Information Processing Systems 35 (NeurIPS)*; Nov 28 to Dec 9, 2022; New Orleans, LA, USA. [doi: [10.5555/3600270.3602070](https://doi.org/10.5555/3600270.3602070)]
173. Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners. *arXiv*. Preprint posted online on May 28, 2020. [doi: [10.48550/arXiv.2005.14165](https://doi.org/10.48550/arXiv.2005.14165)]
174. Fu Y, Peng H, Sabharwal A, Clark P, Khot T. Complexity-based prompting for multi-step reasoning. Presented at: *Proceedings of the Eleventh International Conference on Learning Representations (ICLR 2023)*; May 1-5, 2023; Kigali, Rwanda.
175. Hu EJ, Shen Y, Wallis P, et al. Low-rank adaptation of large language models. Presented at: *Proceedings of the Tenth International Conference on Learning Representations (ICLR 2022)*; Apr 25-29, 2022. [doi: [10.48550/arXiv.2106.09685](https://doi.org/10.48550/arXiv.2106.09685)]
176. Zhang T, Kishore V, Wu F, Weinberger KQ, Artzi Y. Bertscore: evaluating text generation with BERT. *arXiv*. Preprint posted online on Apr 21, 2019. [doi: [10.48550/arXiv.1904.09675](https://doi.org/10.48550/arXiv.1904.09675)]
177. Banerjee S, Lavie A. METEOR: an automatic metric for MT evaluation with improved correlation with human judgments. Presented at: *The ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*; Jun 29, 2005; Ann Arbor, MI. URL: <https://aclanthology.org/W05-0909/> [Accessed 2026-06-19]
178. Ondov B, Attal K, Demner-Fushman D. A survey of automated methods for biomedical text simplification. *J Am Med Inform Assoc*. Oct 7, 2022;29(11):1976-1988. [doi: [10.1093/jamia/ocac149](https://doi.org/10.1093/jamia/ocac149)] [Medline: [36083212](https://pubmed.ncbi.nlm.nih.gov/36083212/)]
179. Hua Y, Na H, Li Z, et al. A scoping review of large language models for generative tasks in mental health care. *NPJ Digit Med*. Apr 30, 2025;8(1):230. [doi: [10.1038/s41746-025-01611-4](https://doi.org/10.1038/s41746-025-01611-4)] [Medline: [40307331](https://pubmed.ncbi.nlm.nih.gov/40307331/)]
180. Freyer O, Wiest IC, Kather JN, Gilbert S. A future role for health applications of large language models depends on regulators enforcing safety standards. *Lancet Digit Health*. Sep 2024;6(9):e662-e672. [doi: [10.1016/S2589-7500\(24\)00124-9](https://doi.org/10.1016/S2589-7500(24)00124-9)] [Medline: [39179311](https://pubmed.ncbi.nlm.nih.gov/39179311/)]
181. Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *arXiv*. Preprint posted online on Apr 12, 2021. [doi: [10.48550/arXiv.2005.11401](https://doi.org/10.48550/arXiv.2005.11401)]
182. Wiest IC, Ferber D, Zhu J, et al. Privacy-preserving large language models for structured medical information retrieval. *NPJ Digit Med*. Sep 20, 2024;7(1):257. [doi: [10.1038/s41746-024-01233-2](https://doi.org/10.1038/s41746-024-01233-2)] [Medline: [39304709](https://pubmed.ncbi.nlm.nih.gov/39304709/)]
183. Alfattni G, Peek N, Nenadic G. Extraction of temporal relations from clinical free text: a systematic review of current approaches. *J Biomed Inform*. Aug 2020;108:103488. [doi: [10.1016/j.jbi.2020.103488](https://doi.org/10.1016/j.jbi.2020.103488)] [Medline: [32673788](https://pubmed.ncbi.nlm.nih.gov/32673788/)]
184. Sullivan PF, Geschwind DH. Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Cell*. Mar 21, 2019;177(1):162-183. [doi: [10.1016/j.cell.2019.01.015](https://doi.org/10.1016/j.cell.2019.01.015)] [Medline: [30901538](https://pubmed.ncbi.nlm.nih.gov/30901538/)]
185. Meyer BM, Rabl U, Huemer J, et al. Prefrontal networks dynamically related to recovery from major depressive disorder: a longitudinal pharmacological fMRI study. *Transl Psychiatry*. Feb 4, 2019;9(1):64. [doi: [10.1038/s41398-019-0395-8](https://doi.org/10.1038/s41398-019-0395-8)] [Medline: [30718459](https://pubmed.ncbi.nlm.nih.gov/30718459/)]
186. Kjaerstad HL, de Siqueira Rotenberg L, Knudsen GM, et al. The longitudinal trajectory of emotion regulation and associated neural activity in patients with bipolar disorder: a prospective fMRI study. *Acta Psychiatr Scand*. Dec 2022;146(6):568-582. [doi: [10.1111/acps.13488](https://doi.org/10.1111/acps.13488)] [Medline: [36054343](https://pubmed.ncbi.nlm.nih.gov/36054343/)]
187. Sinha PP, Mishra R, Sawhney R, Mahata D, Shah RR, Liu H. #suicidal - a multipronged approach to identify and explore suicidal ideation in twitter. Presented at: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM '19)*; Nov 3-7, 2019; Beijing China. [doi: [10.1145/3357384.3358060](https://doi.org/10.1145/3357384.3358060)]

## Abbreviations

**ASD:** autism spectrum disorder

**BERT:** CM Encoder Representations from Transformers

**COT:** chain-of-thought

**DL:** deep learning

**DSM-V:** *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*

**EHR:** electronic health record

**HL7 FHIR:** Health Level 7 Fast Healthcare Interoperability Resources

**ICD-9:** *International Classification of Diseases, Ninth Revision*

**ICD-10:** *International Classification of Diseases, Tenth Revision*

**IE:** information extraction

**LDA:** latent Dirichlet allocation

**LLM:** large language model

**LoRA:** low-rank adaptation

**ML:** machine learning

**NLP:** natural language processing

**OMOP:** observational medical outcomes partnership

**PLM:** pretrained language model

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

**SNOMED-CT:** Systematized Nomenclature of Medicine–Clinical Terms

**TC:** text classification

**UMLS:** Unified Medical Language System

---

*Edited by Arriel Benis; peer-reviewed by K Pradeep, Md Muntasir Zitu; submitted 12.Jan.2026; final revised version received 21.Apr.2026; accepted 11.Jun.2026; published 10.Jul.2026*

*Please cite as:*

Rao S, Chen Xng, Deng G, Xie J, Jiang T, Li T, Zhang Y, Jiang H

*Natural Language Processing Applied to Psychiatric Clinical Notes: Scoping Review*

*JMIR Med Inform* 2026;14:e91249

URL: <https://medinform.jmir.org/2026/1/e91249>

doi: [10.2196/91249](https://doi.org/10.2196/91249)

© Shuying Rao, Xi'ang Chen, Guifeng Deng, Junyi Xie, Tiecheng Jiang, Tao Li, Yaoyun Zhang, Haiteng Jiang. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 10.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org>, as well as this copyright and license information must be included.