International Platform of Registered Systematic Review and Meta-analysis Protocols

INPLASY

Machine Learning for the Detection of Mental Health Disorders Using Clinical Text: A Systematic Review

INPLASY202650010
doi: 10.37766/inplasy2026.5.0010
Received: 4 May 2026
Published: 4 May 2026

Ding, XC; Lam, CLM; Yim, SH; Cheung, AK; Lee, TMC.

Corresponding author:
Xiaochen Ding

xcding@hku.hk

Author Affiliation:
InnoCentre of Clinical Neuropsychology, The University of Hong Kong, Hong Kong SAR, China.

ADMINISTRATIVE INFORMATION

Support - None.

Review Stage at time of this submission - Data extraction.

Conflicts of interest - None declared.

INPLASY registration number: INPLASY202650010

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 4 May 2026 and was last updated on 4 May 2026.

INTRODUCTION

Review question / Objective The primary objective of this systematic review is to evaluate the classification performance of machine learning and natural language processing models in detecting formally diagnosed mental health disorders using unstructured real-life clinical text.

Rationale This research fosters integration of traditional diagnostics with AI technologies, establishing scientific foundations for building more precise and efficient mental healthcare systems. By highlighting current technological limitations and future directions, the study offers practical guidance for the global digital transformation of mental health services, particularly for enhancing service accessibility and ensuring diagnostic equity, while proposing actionable solutions.

Condition being studied The conditions being studied encompass any specific mental health disorder officially recognized in the DSM-5. Based on the DSM-5 Table of Contents, this broadly includes, but is not limited to, depressive disorders, anxiety disorders, schizophrenia spectrum and other psychotic disorders, bipolar and related disorders, trauma- and stressor-related disorders, neurodevelopmental disorders, eating disorders, and substance-related and addictive disorders.

METHODS

Participant or population Patients receiving care in formal healthcare or clinical settings. The population includes individuals whose health data is formally documented in electronic health records, electronic medical records, clinical notes, discharge summaries, transcribed clinical interviews, or other official clinical narratives. Studies focusing on populations from non-clinical settings will be strictly excluded.

Intervention The application of machine learning, deep learning, artificial intelligence, or natural

language processing models (including large language models) designed to detect, predict, or classify mental health disorders. The models must actively rely on unstructured clinical text data as a primary input for their predictive development and function.

Comparator The comparator is the formal clinical diagnosis or documented psychiatric diagnostic criteria extracted directly from the patient's health record by healthcare professionals.

Study designs to be included Peer-reviewed journal articles and conference papers with completed experimental research.

Eligibility criteria Studies must be published between 2016 and 2026 and restricted to the English language. Studies will be excluded if they rely entirely or partially on non-clinical text, rely exclusively on non-textual data without applying NLP to clinical text, or have absent, incomplete, or purely qualitative reporting of the model's performance metrics.

Information sources A comprehensive systematic search was conducted on April 7, 2026, across the following five electronic databases: IEEE Xplore, Web of Science, PubMed, Scopus, and PsycINFO.

Main outcome(s) The primary outcome is the quantitative statistical report of the machine learning model's classification performance in detecting or predicting mental health disorders. This includes standard classification metrics such as Accuracy, Sensitivity (Recall), Specificity, Precision, F1-score, and AUC-ROC.

Data management References retrieved from the literature searches will be exported to EndNote 2025 for automated initial deduplication. The remaining records will then be imported into Rayyan for title and abstract screening, as well as a secondary manual deduplication process. Two independent reviewers will screen the records, and conflicts will be resolved through discussion. Full-text screening and final data extraction will be managed and recorded using a structured spreadsheet.

Quality assessment / Risk of bias analysis The risk of bias and concerns regarding applicability for the included studies will be assessed using the PROBAST (Prediction model Risk Of Bias ASsessment Tool) framework.

Strategy of data synthesis A narrative synthesis and tabular presentation of the extracted data will be conducted. We will systematically summarize the characteristics of the included studies, focusing on the specific mental health disorders targeted, the types of clinical text utilized, and the specific NLP or machine learning architectures employed. The primary synthesis will compare the classification performance metrics across the different models and target disorders.

Subgroup analysis If the extracted data permits, narrative subgroup analyses will be conducted to explore variations in model performance based on the specific category of mental health disorder targeted and the type of machine learning or NLP architecture employed.

Sensitivity analysis As a formal quantitative meta-analysis is not anticipated due to expected heterogeneity, a narrative sensitivity analysis will be conducted.

Language restriction English.

Country(ies) involved Hong Kong SAR, China.

Keywords Machine Learning; Natural Language Processing; Mental Health Disorders; Electronic Health Records; Clinical Text; Diagnostic Classification.

Contributions of each author
Author 1 - Xiaochen Ding.
Author 2 - Charlene L.M. Lam.
Author 3 - See Heng Yim.
Author 4 - Amanda K. Cheung.
Author 5 - Tatia M.C. Lee.