

Original Paper

AI Models for Suicide Risk Prediction in Adult Patients Receiving Mental Health Care Using Real-World Data: Retrospective Population-Based Study

Cleofé Peña-Gómez^{1,2*}, PhD; Marc Fradera^{2,3,4*}, PhD; Xavier Sánchez Corrales⁵, MS; Marc Caravaca-Rodriguez², BSc; Juan-Francisco Martínez-Cerdá⁶, PhD; Joan Albert Escofet⁶, MS; David Roche⁷, PhD; Eneko Barbería^{8,9}, MD; Caridad Pontes^{10,11,12}, PhD, MD; Jesús Giraldo^{3,4,13}, PhD; Diego Palao^{2,3,4,5,14}, MD, PhD

¹Barcelona Supercomputing Center, Barcelona, Spain

²Institut d'Investigació i Innovació Parc Taulí (I3PT-CERCA), Sabadell, Spain

³Unitat Mixta de Neurociències Traslacional I3PT-INc-UAB, Universitat Autònoma de Barcelona, 08208 Sabadell, Spain

⁴Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Instituto de Salud Carlos III, 28029 Madrid, Spain

⁵Department of Mental Health, University Hospital Parc Taulí, 08208 Sabadell, Spain

⁶Agency for Health Quality and Assessment of Catalonia (AQUAS), Data Analytics Program for Health Research and Innovation (PADRIS) of the Data and AI Department, Barcelona, Spain

⁷Research Institute for Evaluation and Public Policies (IRAPP), Universitat Internacional de Catalunya (UIC), Barcelona, Spain

⁸Forensic Pathology Department, Institut de Medicina Legal i Ciències Forenses de Catalunya (IMLCFC), Barcelona, Spain

⁹School of Medicine and Health Sciences, Universitat Rovira i Virgili (URV), Reus, Spain

¹⁰Digitalization for the Sustainability of the Healthcare System (DS3), Barcelona, Spain

¹¹Servei de Farmacologia Clínica, Hospital de la Santa Creu i de Sant Pau, Barcelona, Spain

¹²Departament de Farmacologia, de Toxicologia i de Terapèutica, Universitat Autònoma de Barcelona, Barcelona, Spain

¹³Laboratory of Molecular Neuropharmacology and Bioinformatics, Unitat de Bioestadística, Institut de Neurociències, Universitat Autònoma de Barcelona, Bellaterra, Spain

¹⁴Department of Psychiatry and Forensic Medicine, Medical School, Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès) Barcelona, Spain

*these authors contributed equally

Corresponding Author:

Cleofé Peña-Gómez, PhD

Institut d'Investigació i Innovació Parc Taulí (I3PT-CERCA)

Plaça Torre de l'Aigua, s/n,

Sabadell 08208

Spain

Phone: +34 937458376

Email: cpenagom@bsc.es

Abstract

Background: Suicidal behavior is a major public health problem worldwide. The exact etiology remains unclear, representing a complex problem involving multiple factors. Evidence indicates that around 50% to 80% of people who die by suicide have had contact with the health care system in the year prior to their death.

Objective: We present the IDICIUS project, whose objective is to develop a clinical decision support system that functions as an early warning system and applies AI to prevent suicide risk using anonymized electronic health record data.

Methods: This study shows the first 2 stages of the IDICIUS project, where real-world data from 4 public sources were integrated, curated, standardized, and anonymized in phase 1, and analysis and modeling of suicide risk were performed in phase 2. This retrospective population-based study included 41,557 adult patients receiving mental health care at a large hospital. We evaluated the performance of machine learning classifiers with increasing complexity: logistic regression, elastic net, decision trees and random forests, bagging and boosting ensemble methods (GradientBoosting, XGBoost [extreme gradient boosting], CatBoost, RandomForest, and AdaBoost [adaptive boosting]), support vector machines, and deep neural networks. To address class imbalance, we tested several balancing techniques, with random undersampling providing the best results.

Results: Phase 1 yielded a useful database of 32,661 patients receiving mental health care and 112 features. Of those, 2764 patients exhibited suicidal behavior (target), while 29,897 did not (control). The undersampling reduced the class distribution to 4422 vs 2211 (control vs target). The most prevalent features in the target class (N=2764) were psychiatric conditions, anxiety episodes (n=2269, 82.1%) and depressive episodes (n=1682, 60.9%), along with demographic and behavioral factors: female sex (n=1672, 60.5%), alcohol consumption (n=770, 27.9%), and conduct disorder (n=709, 25.7%). The largest patient subgroups were females with combined depression and anxiety (n=300, 10.9%), females with anxiety disorders only (n=182, 6.6%), males with anxiety disorders only (n=120, 4.3%), and males with combined depression and anxiety (n=106, 3.8%). In phase 2, ensemble methods, especially GradientBoosting and XGBoost, achieved the best performance, with receiver operating characteristic-area under the curve scores around 0.95. While these models showed moderate precision in identifying true positives (0.51-0.58), they demonstrated high sensitivity in detecting at-risk patients, with recall scores of 0.85 and 0.84, respectively. Both models achieved a good balance between precision and recall (F_1 -scores: 0.68 and 0.67). The strongest predictors were corticosteroid use, psychiatric medications (olanzapine and lorazepam), female sex, and depression.

Conclusions: We integrated multiple electronic health record data sources and applied AI to develop, model, and optimize ensemble algorithms for suicide prevention, maximizing effectiveness, efficiency, and generalizability. These AI-generated algorithms, based on readily available and well-structured data, may advance the identification of at-risk patients and function as early warning systems, thereby increasing opportunities for timely preventive interventions.

JMIR AI 2026;5:e85085; doi: [10.2196/85085](https://doi.org/10.2196/85085)

Keywords: suicide prevention; suicidal behavior; mental health; real-world data; clinical decision support system; early warning system; electronic health records; AI; machine learning

Introduction

Suicidal behavior represents a complex problem with a substantial impact on public health worldwide [1-3]. Its heterogeneity and multifactorial nature make prediction in clinical practice particularly challenging, especially in the medium and long terms [4]. Nevertheless, evidence indicates that many individuals who die by suicide have sought some form of help prior to their death [5]. Approximately 50% to 80% of people who died by suicide had a prior interaction with the public health care system. Of these, around 60% did so during the previous month [6,7].

The implementation of early-stage interventions in this context is particularly crucial, as a prompt identification of individuals at high and imminent risk may allow for the effective mitigation of progression along the suicidal behavior continuum before it advances to more severe phases [8,9].

Although most prevention programs rely on training health care professionals to detect potential suicide risk and address it, recent reviews and meta-analyses suggest that such preventive measures may not achieve the expected effectiveness [10-14], and clinicians generally perform poorly at predicting who will attempt suicide [15]. This limitation highlights the need for novel approaches to identify patients at risk within health care settings—approaches designed not to replace clinical judgment but to complement and enhance existing assessment strategies [16].

In this context, electronic health records (EHRs) typically include longitudinal information, allowing for the monitoring of patients over extended periods and the identification of temporal patterns preceding suicidal behavior [17]. EHRs integrate a wide range of variables, including sociodemographic characteristics, clinical diagnoses, treatment histories, comorbidities, and health care utilization patterns [7]. Multiple studies have demonstrated that applying

AI-based tools to EHR data can identify novel characteristics beyond the well-established risk factors for suicidal behavior [17-22], thereby enhancing the predictive accuracy of suicide risk models compared with previous approaches [17, 23,24]. Furthermore, AI approaches using machine learning (ML) models enable the simultaneous assessment of multiple variables, capturing complex interdependencies [25, 26] and nonlinear associations [27,28] in the development of predictive models.

The use of AI has also been extended to the development of early warning systems (EWSs) in hospitals as a clinical decision support system (CDSS) for clinicians [29, 30]. These systems typically generate risk scores designed to assist clinicians in recognizing clinical deterioration among hospitalized patients. When integrated into EHRs, some EWSs have been associated with reductions in in-hospital mortality [31,32]. Building on this concept, similar AI-driven early warning approaches could be adapted to mental health care, enabling the timely identification of individuals at elevated risk for suicide and facilitating prompt, targeted interventions.

The IDICIUS project aims to develop a CDSS, motivated by EWS, that applies AI predictive models to prevent suicide risk. This tool will integrate information on suicide attempts, suicidal ideation, and deaths by suicide with clinical and sociodemographic factors derived from real-world data (RWD) recorded in EHRs and public health care and/or administrative databases, based on a cohort of 41,557 adult patients receiving psychiatric care. The IDICIUS project consists of a 4-stage approach, including (1) gathering of usable data, (2) analysis and modeling of risk, (3) building of a predictive tool, and (4) the implementation of the tool for effectiveness assessment. This study presents the first 2 stages of the IDICIUS project. Phase 1 focuses on integrating and harmonizing multiple databases with input from relevant stakeholders; implementing data curation, standardization,

and anonymization; and defining key risk factors. Phase 2 aims to develop and validate predictive models using advanced ML techniques, selecting and optimizing the most effective, efficient, and generalizable algorithm for suicide prevention. Ultimately, these models are designed to assist clinicians in monitoring variations in suicide risk among patients and to be implemented within routine clinical workflows.

Methods

Sample Characteristics and Study Design

This is a retrospective, population-based study that used structured and anonymized existing EHR data from a large cohort of prevalent adult patients receiving mental health follow-up care in a large hospital facility (Consorci Corporació Sanitària Parc Taulí). The hospital provides outpatient care to a population of approximately 440,000 inhabitants in Sabadell, Catalonia. Predictive models were developed using AI algorithms and ML techniques, trained on the obtained data.

The study included data from cases who received mental health care between January 1, 2018, and June 1, 2024, at the hospital. The index visit was defined as the first recorded clinical encounter that qualified for study inclusion. To be eligible, patients were required to have had at least one additional visit within the 3 months preceding this index visit. Exclusion criteria included nonresidency in the Province of Barcelona and the absence of any visit to a primary care health center.

Ethical Considerations

The study received approval from the institutional ethics review board under study code TED2021-131979B-C31 and

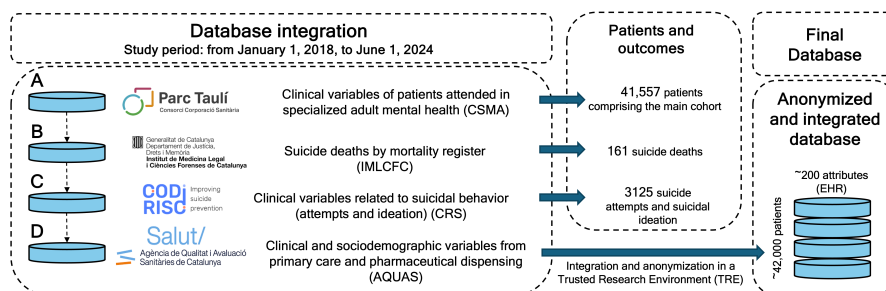
case number 2023/5120. In addition, the study was approved by the institutional ethics committee to ensure compliance with the General Data Protection Regulation and applicable laws. As it used existing, fully anonymized data processed by an independent team, informed consent was waived by the ethics committee. Data confidentiality was maintained in accordance with Spanish Organic Law 3/2018 [33] and Regulation (EU) 2016/679 [34]. The study was conducted in accordance with the principles of the Declaration of Helsinki. The review process included 2 rounds of evaluation, with the initial submission on November 28, 2023, and final approval granted on January 25, 2024.

RWD Sources

RWD was linked and extracted from 4 primary data sources (Figure 1):

1. EHRs of patients in mental health follow-up, attended at any level of care within the Parc Taulí Mental Health Service during the study period.
2. Catalonia Suicide Risk Code registry, containing sociodemographic and clinical information for individuals who have attempted suicide and/or present suicidal ideation assessed as high risk.
3. Registry of suicide deaths from the Institut de Medicina Legal i Ciències Forenses, the official source for identifying deaths by suicide.
4. Outpatient medication database of prescriptions billed by pharmacies to the Catalan Health Service, comprising all medications prescribed through the public health care system, dispensed by community pharmacies, and reimbursed.

Figure 1. Real-world data integration schema. Four sources of information were integrated: (A) electronic health records (EHRs) registered from January 1, 2018, to June 1, 2024, at specialized mental health units for adults (Centre de Salut Mental per Adults [CSMA]) at Parc Taulí University Hospital (Consorci Corporació Sanitària Parc Taulí [CCSPT]); (B) deaths by suicide recorded by the forensic institute (Institut de Medicina Legal i Ciències Forenses de Catalunya [IMLFCF]); (C) clinical variables specifically related to suicidal behavior (Codi Risc Suïcidi [CRS]) registered by the Department of Health of Catalonia; and (D) clinical, sociodemographic variables, and pharmacy dispensation data registered by the Catalan Health Service (Catsalut) for invoicing purposes. All data were linked, anonymized, processed, and provided by AQUAS (Agència de Qualitat i Avaluació Sanitàries de Catalunya), a public-law entity under the Department of Health of Catalonia. After initial harmonization, data curation, and integration, performed at the Trusted Research Environment (TRE), a database of 41,557 patients and 206 features was obtained.



Given the highly sensitive nature of the data, data linkage and extraction were performed by a team independent of the researchers, in accordance with legal requirements. The 4 data sources were subsequently harmonized, integrated,

and anonymized by personnel from the Agency for Health Quality and Assessment of Catalonia (AQUAS [Agència de Qualitat i Avaluació Sanitàries de Catalunya]), a public-law entity under the Department of Health of the Generalitat

of Catalonia holding the custody of large public health care databases for academic research. All data processing and modeling were conducted within AQUAS's Trusted Research Environment, a secure computational infrastructure that prevents extraction of individual-level microdata under any circumstances and allows access to anonymized data only to authorized study personnel designated by the Department of Health of the Generalitat of Catalonia. Only aggregated statistical outputs were permitted for export, and these required explicit prior authorization through established data governance protocols.

ML Classifiers and Hyperparameter Tuning

We evaluated the performance of state-of-the-art ML classifiers with an increasing degree of complexity: logistic regression, elastic net, decision trees and random forests, bagging and boosting ensemble methods (GradientBoosting, extreme gradient boosting [XGBoost], CatBoost, RandomForest, and adaptive boosting [AdaBoost]), support vector machines, and deep learning (DL). For the DL approach, we implemented a sequential neural network with 3 hidden layers (128, 64, and 8 units) using rectified linear unit activation, L2 regularization (0.001), and progressively adjusted dropout rates (0.3, 0.4, and 0.5). The output layer used a sigmoid activation for prediction. EarlyStopping and ReduceLROnPlateau were applied to prevent overfitting.

The optimal hyperparameters for the trained models were determined through grid search combined with 10-fold cross-validation, enabling systematic and robust tuning within the algorithms' search space.

Model Performance

Model performance on the target class was evaluated using standard performance metrics precision, recall, accuracy, and the F_1 -score, all of which are derived from true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) outcomes. Precision ($TP/[TP + FP]$) reflects how reliable the identified risk cases are; recall ($TP/[TP + FN]$) measures how many at-risk patients were successfully detected—that is, the ability to identify positives; accuracy represents the overall model correctness ($(TP + TN)/[TP + FN + TN + FP]$); and the F_1 -score combines precision and recall to measure how well the model balances correctly identifying patients who have the condition while avoiding misdiagnosing patients who do not, by using the harmonic mean: $(2 \times \text{precision} \times \text{recall})/(\text{precision} + \text{recall})$.

In our clinical context, recall is the most critical metric, as an FN (failing to identify a patient at suicidal risk) can have severe consequences. Therefore, the model was designed to prioritize minimizing FNs, even at the expense of generating more FPs, which can subsequently be ruled out through clinical evaluation.

Dataset Characteristics and Preparation

The integrated full dataset, combining the 4 RWD sources, consisted of 41,557 eligible patients with a total of 206 features. To construct the analysis dataset, one patient with

a response rate below 5% was excluded, and the age range was restricted to 18 to 95 years to focus the study on the adult population, as risk factors significantly differ between adult and pediatric populations. The variable that registered the suicide attempts in the past 30 days ("count_temp_30d") was also excluded to prevent data leakage with the target variable. The resulting analysis dataset comprised 32,661 patients and 112 features, ensuring data quality and relevance for subsequent analyses. Of those, 2764 patients exhibited suicidal behavior (target class), while 29,897 did not (control class).

Feature intersection analysis was performed using UpSet plots to visualize the complex overlapping patterns among features associated with the target class. This approach allows for the comprehensive examination of multiway intersections that extend beyond the limitations of traditional Venn diagrams. The analysis identified all possible combinations of these features and quantified the size of each intersection within the study population.

The analysis dataset was split into 80% training and 20% testing sets. To address class imbalance, we evaluated several balancing techniques, including SMOTE (Synthetic Minority Oversampling Technique), ADASYN (Adaptive Synthetic Sampling), SMOTE-Tomek, the Neighborhood Cleaning Rule, and RandomUnderSampler, both individually and in combination. Among these approaches, the highest performance was achieved using RandomUnderSampler alone. The RandomUnderSampler (sampler_strategy=0.5) reduced the class distribution from 23,917 vs 2211 to 4422 vs 2211 (negative vs positive cases). Prior to feeding the models, standard scaling was applied only to the numerical features (a total of 21).

To enhance model performance, we conducted additional feature engineering by applying log transformations to numerical variables and testing various mathematical transformations (divisions, sums, and logarithms) on key variables, including age, number of diagnoses, and medication count. However, these modifications did not improve the final model metrics.

Explainability

We computed SHAP [35] (Shapley additive explanations) values for the classifiers to assess feature importance. The summary plot ranks features by influence (Y axis), while the X axis shows each feature's impact on predictions. Positive SHAP values increase predicted suicide risk, whereas negative values decrease it, with color indicating high (red) or low (blue) feature values.

Results

Sample and Dataset Characteristics

The study included 41,557 cases and 206 features in the full dataset and 32,661 patients and 112 features in the analysis dataset. During the study period, 2604 individuals had a recorded suicide attempt and 160 died by suicide. In total,

2764 individuals had suicidal behavior and were defined as the target class (Table 1).

Among the target class (N=2764), the most prevalent features were psychiatric conditions, including anxiety episodes (n=2269, 82.1%) and depressive episodes (n=1682, 60.9%), followed by female sex (n=1672, 60.5%), alcohol consumption (n=770, 27.9%), and conduct disorder (n=709, 25.7%; Multimedia Appendix 1). The most common

subgroup was females with both depression and anxiety (n=300, 10.9%), followed by females with anxiety disorders (n=182, 6.6%), males with anxiety disorders (n=120, 4.3%), and males with both depression and anxiety (n=106, 3.8%). In the analysis dataset, 60.5% (1672/2764) of suicidal behavior cases occurred in women; the prevalence of suicidal behavior was 9.0% (1672/32,662) in women and 7.7% (1092/32,662) in men.

Table 1. Sociodemographic characteristics and distribution of suicidal behavior.

Characteristic	Full dataset	Analysis dataset
Sample (n)	41,557	32,662
Age (y), mean (SD; range)	39.57 (22.19; 1-118)	47.08 (19.39; 18-95)
Sex distribution (total sample)		
Female, n (%)	22,150 (53.3)	18,518 (56.7)
Male, n (%)	19,407 (46.7)	14,144 (43.3)
Sex distribution by suicidal behavior		
No suicidal behavior, n (%)		
Female, n (%)	38,432 (92.5)	29,898 (91.5)
Male, n (%)	20,182 (52.5)	16,846 (56.3)
Suicidal behavior, n (%)		
Female, n (%)	18,250 (47.5)	13,052 (43.7)
Male, n (%)	3125 (7.5)	2764 (8.5)
Suicidal behavior rates by sex		
Female, n (%)	1968 (8.9)	1672 (9)
Male, n (%)	1157 (6)	1092 (7.7)

Model Performance

Among the evaluated models, the ensemble methods GradientBoosting, XGBoost, CatBoost, RandomForest, and AdaBoost ranked highest, with receiver operating characteristic-area under the curve (ROC-AUC) values

between 0.94 and 0.95 (Table 2 and Figure 2). These models also showed the highest accuracy, all above 0.91. Their precision ranged from 0.51 to 0.58, and recall ranged from 0.82 to 0.86. Support vector machine had the lowest precision (0.35) and recall of 0.81.

Figure 2. Receiver operating characteristic (ROC) curves for all models. ROC curves comparing the performance of 11 machine learning models for suicide risk prediction. GradientBoosting achieved the highest area under the curve (AUC; 0.958), followed closely by XGBoost (0.958) and CatBoost (0.957). Tree-based ensemble methods consistently outperformed traditional machine learning approaches, with LogisticRegression and ElasticNet showing the lowest performance (AUC=0.844 and 0.842, respectively). AdaBoost: adaptive boosting; ANN: artificial neural network; KNN: k-nearest neighbor; SVM: support vector machine; XGBoost: extreme gradient boosting.

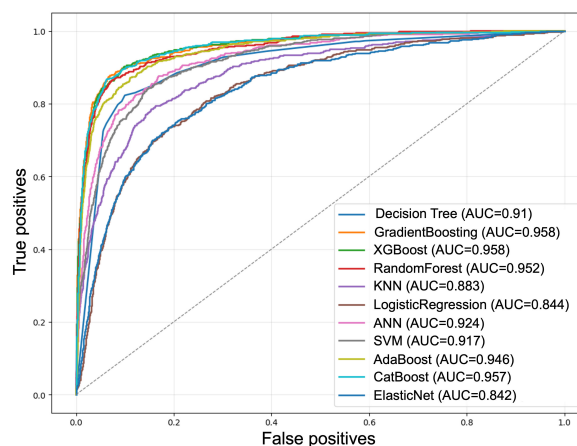


Table 2. Model performance metrics.

Model	ROC-AUC ^a	Accuracy	F_1 -score	Precision	Recall
GradientBoosting	0.9582	0.9343	0.6889	0.5751	0.8590
XGBoost ^b	0.9581	0.9313	0.6763	0.5624	0.8481
CatBoost	0.9574	0.9348	0.6872	0.5785	0.8463
RandomForest	0.9521	0.9300	0.6705	0.5576	0.8409
AdaBoost ^c	0.9458	0.9172	0.6256	0.5067	0.8174
ANN ^d	0.9241	0.9092	0.5850	0.4772	0.7559
SVM ^e	0.9165	0.8609	0.4975	0.3583	0.8137
DecisionTree	0.9104	0.8938	0.5673	0.4329	0.8228
KNN ^f	0.8826	0.9105	0.5193	0.4759	0.5714
LogisticRegression	0.8436	0.8054	0.3881	0.2644	0.7288
ElasticNet	0.8419	0.7964	0.3814	0.2567	0.7414

^aROC-AUC: receiver operating characteristic-area under the curve.

^bXGBoost: extreme gradient boosting.

^cAdaBoost: adaptive boosting.

^dANN: artificial neural network.

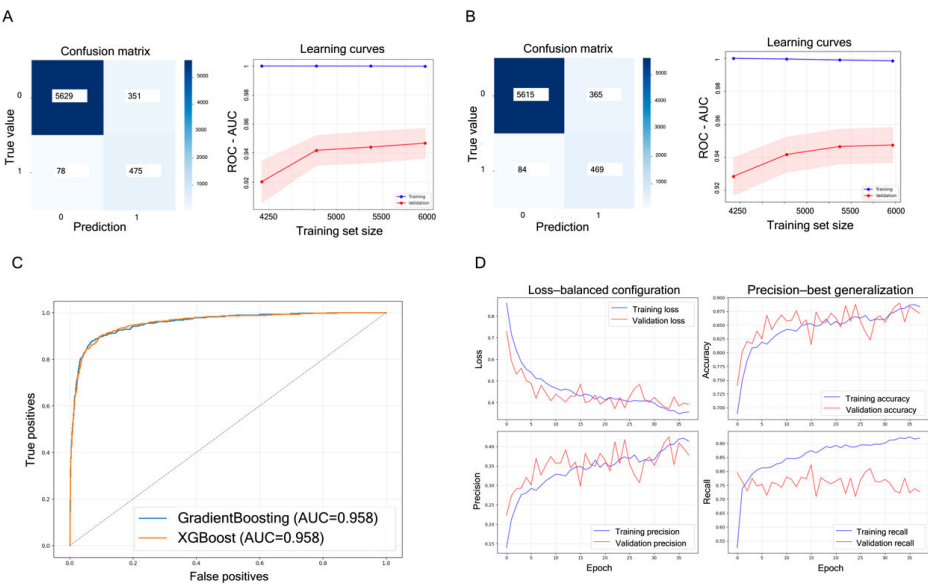
^eSVM: support vector machine.

^fKNN: k-nearest neighbor.

F_1 -scores varied across models, with the top-performing models ranging from 0.63 to 0.69. GradientBoosting achieved the highest F_1 -score (0.6889), followed by CatBoost (0.6872) and XGBoost (0.6763). Traditional statistical models performed worse across all 5 metrics. LogisticRegression

and ElasticNet had ROC-AUC values below 0.85 and the lowest F_1 -scores (0.38 for both). KNN showed high accuracy (0.91) but lower recall (0.57). Overall, GradientBoosting and XGBoost were the best-performing models (Figure 3).

Figure 3. Performance analysis of top-performing ensemble models. Comprehensive evaluation of (A) GradientBoosting and (B) extreme gradient boosting (XGBoost) models showing confusion matrices and learning curves. The confusion matrices demonstrate high true-negative rates (5629/6533, 86.1% and 5615/6533, 85.9% for GradientBoosting and XGBoost, respectively) with relatively low false-positive and false-negative rates. Learning curves indicate stable performance across training set sizes with minimal overfitting. (C) Direct comparison of receiver operating characteristic (ROC) curves for both ensemble models demonstrates nearly identical performance with overlapping curves and identical area under the curve (AUC) values of 0.95. (D) Neural network training dynamics showing loss and performance metrics over epochs. The balanced configuration demonstrates stable convergence with training and validation losses decreasing consistently, while accuracy, precision, and recall metrics show steady improvement with minimal overfitting, indicating robust model generalization.



To test model stability and metric variability across imbalance-handling methods, we ran 1000 iterations of the best-performing model, GradientBoosting, under each

strategy and summarized performance using mean estimates and 95% CIs (Table 3).

Table 3. Mean performance and 95% CIs of GradientBoosting across 1000 bootstrap resamples under different class-balancing methods.

Class-balancing method	ROC-AUC ^a , mean (95% CI)	Accuracy, mean (95% CI)	F ₁ -score, mean (95% CI)	Precision, mean (95% CI)	Recall, mean (95% CI)
RandomUnderSampler	0.958 (0.946-0.969)	0.934 (0.926-0.942)	0.689 (0.648-0.726)	0.575 (0.527-0.622)	0.859 (0.816-0.898)
SMOTE ^b	0.924 (0.907-0.939)	0.919 (0.909-0.928)	0.615 (0.570-0.657)	0.514 (0.463-0.566)	0.765 (0.714-0.813)
ADASYN ^c	0.920 (0.903-0.936)	0.921 (0.911-0.929)	0.619 (0.577-0.660)	0.523 (0.479-0.569)	0.758 (0.709-0.810)
SMOTE-Tomek	0.921 (0.904-0.937)	0.917 (0.907-0.926)	0.610 (0.567-0.650)	0.506 (0.458-0.552)	0.767 (0.717-0.814)
Cost-sensitive learning (manual weights)	0.953 (0.936-0.968)	0.958 (0.951-0.965)	0.743 (0.69-0.785)	0.782 (0.732-0.828)	0.708 (0.654-0.762)
Cost-sensitive learning (automatic weights)	0.952 (0.936-0.967)	0.958 (0.951-0.965)	0.742 (0.699-0.784)	0.782 (0.726-0.832)	0.707 (0.650-0.758)

^aROC-AUC: receiver operating characteristic-area under the curve.

^bSMOTE: Synthetic Minority Oversampling Technique.

^cADASYN: Adaptive Synthetic Sampling.

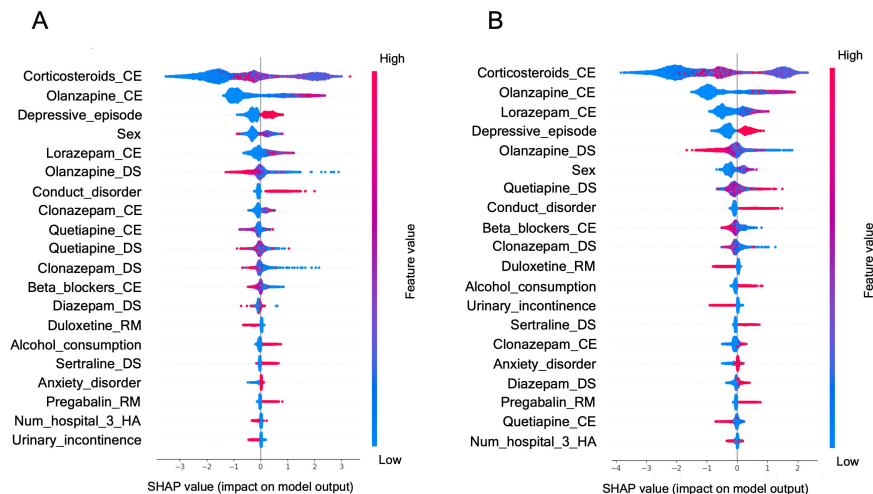
The neural network was evaluated with and without transformed numerical variables, with no relevant differences between the settings. Using the default threshold of 0.5, the model achieved an accuracy of 0.89, a precision of 0.43, a recall of 0.69, and an F_1 -score of 0.53. Threshold optimization identified 0.77 as the optimal cutoff, yielding a precision of 0.56, a recall of 0.58, and an F_1 -score of 0.57.

Explainability

SHAP analysis showed similar feature-importance patterns for GradientBoosting and XGBoost (Figure 4). The most

influential features included corticosteroid use, psychiatric medications such as olanzapine and lorazepam, female sex, and recurrent depression. Features such as urinary incontinence, hospitalizations, pregabalin use, and anxiety disorder were less influential.

Figure 4. SHAP (Shapley additive explanations) values showing the impact of top features on suicide risk prediction for (A) GradientBoosting and (B) extreme gradient boosting (XGBoost) models. The consistent patterns between models validate the robustness of these feature importance rankings. Each point represents a patient, with the X axis indicating the feature’s contribution to the model’s output. Positive values increase suicide risk prediction, while negative values decrease it. Colors represent feature values (red color denotes “high” and blue color denotes “low”). Suffixes indicate drug intake timing: CE: current; DS: distant; RM: remote.



Discussion

This study evaluates the potential of AI for suicide prevention in a psychiatric setting using harmonized clinical data from 4 public databases. In a cohort of 32,661 patients

receiving mental health care, suicidal behavior was associated with a relatively small set of common clinical and pharmacological features, including anxiety and depressive episodes, conduct disorders, corticosteroids, lorazepam, clonazepam, and olanzapine. Ensemble methods outperformed traditional

approaches, in line with prior evidence that boosting methods often perform strongly across settings [36].

GradientBoosting and XGBoost showed the best overall performance (F_1 -scores of 0.68 and 0.67, respectively). In screening or triage settings, where FNs are particularly consequential, recall should be as high as possible. Here, recall was relatively high (0.85 and 0.84, respectively), supporting the potential use of these models alongside clinical judgment to guide targeted interventions. Accordingly, threshold selection should be guided by the clinical context and the relative costs of FPs and FNs [37]. A previous meta-analysis assessing ML performance in clinical populations reported a pooled recall of 0.58 [38]. Similarly, Pigoni et al [39] found substantial heterogeneity in recall/sensitivity across ML studies in psychiatric clinical populations, with values ranging from 0.13 to 0.98 and an overall mean of 0.72 after harmonizing percentages and range estimates. Traditional clinical judgment and standardized risk scales have demonstrated only modest predictive performance, slightly better than chance, although they may capture contextual and behavioral cues not reflected in structured data, such as affect or tone of voice [4,15]. At the same time, a substantial proportion of individuals who die by suicide have recent contact with health services, up to 80% within the year prior and 50% within the month before death [7]. Together, these findings highlight the potential value of data-driven approaches to support the identification of risk signals that may not be consistently detected during routine clinical assessment.

Precision was moderate (0.56–0.57), indicating that some flagged individuals may not ultimately be at high risk. In the context of psychiatric follow-up, the additional burden of risk flags may be more manageable, particularly when alerts are restricted to patients already receiving care and when the output is used to support rather than replace clinical judgment. Although FPs can still contribute to unnecessary workload, the potential benefits of preventing suicide may justify a low-to-moderate positive predictive value when the intervention is cost-effective, actionable, and integrated into routine care. Indeed, comparable low positive predictive value strategies are routinely employed in other areas of health care, including breast cancer screening and stroke prevention [40]. Nevertheless, FPs should be minimized through threshold calibration, tiered risk stratification, and workflow integration.

Previous studies applying AI to EHRs have yielded mixed findings, with some reporting strong predictive power [27] and others showing good performance with classical statistical models such as LASSO or logistic regression [19]. Such variability likely reflects differences in methods, sample sizes, populations, and feature selection. Importantly, we excluded variables capturing prior suicide attempts, regardless of their known clinical value, to avoid data leakage. Although including these features markedly improved performance (precision from 0.56 to 0.98 for boosting methods), the models became overly dependent on them. Instead, we aimed to develop a clinically useful EWS that prioritizes recall and can identify individuals at risk even

in the absence of prior attempts—when prevention is most challenging yet most critical.

Corticosteroid use consistently ranked as a highly important feature across both models, suggesting that it captures key clinical information, possibly reflecting chronic and severe underlying physical conditions linked to chronic suffering, underlying inflammatory conditions, or even a risk for corticosteroid-induced psychiatric symptoms and behavioral disturbances, and relapse or worsening of underlying psychiatric conditions [41]. Its prominence, alongside any other medications, may also indicate broader systemic issues. In GradientBoosting, corticosteroid use showed high variability in impact (SHAP values ~ -2 to $+3$), while in XGBoost, its effect was similarly important but more constrained. Antipsychotic medications (olanzapine and quetiapine) and dosing patterns also contributed notably, potentially reflecting their role as markers for underlying severe psychiatric morbidity, such as schizophrenia, impulsivity disorders, or complicated depression, which are usual indications for their use and may be the actual underlying risk factors [7]. Similarly, clonazepam is often used for insomnia, a clinical condition that has been reported to increase the risk of suicide [42]. In these cases, exposure to these medications may be a useful marker of risk and is convenient for algorithms, since these are structured data that are readily available in databases [43]. Benzodiazepines, such as lorazepam and diazepam, ranked as a highly important feature across both models, are commonly used in psychiatric practice for anxiety, insomnia, and agitation [7]. The models incorporate demographic characteristics (eg, biological sex) alongside behavioral risk indicators (eg, alcohol consumption). Together, these clinically familiar features provide face validity to the model and integrate multiple variables simultaneously rather than relying on any single risk factor, potentially highlighting risk patterns that may not be readily apparent in routine practice. Thus, value is added by ensuring completeness in recovering and summarizing relevant risk information from existing records, easing the health care professional's task. This may be particularly relevant in primary care and emergency settings, where EHRs can be incomplete, fragmented across care environments, or not immediately accessible and where timely risk identification and stratification are essential.

While EHRs provide valuable clinical information, they fail to capture the full multidimensional complexity of individuals' lives. We previously reported [44] that DL models improved the detection of TPs by incorporating domains absent from EHRs, such as risk-taking behavior, social media activity, social connectedness, sexual behavior, and personality traits. In that study, recall ranged from 0.67 to 0.87 and accuracy from 0.83 to 0.87, although precision remained lower. Subsequent work further improved performance, with the best models achieving area under the curve values of 0.85 to 0.90 and recall values of 0.81 to 0.93, but with very low precision of 0.17 to 0.16. These results highlight that suicide risk prediction may benefit from integrating routinely collected clinical data with broader psychosocial and behavioral dimensions.

However, ensuring responsible data integration requires coordinated efforts across stakeholders while maintaining privacy and ethical standards. In addition, generalizability remains a major challenge, as models developed in one population may not transfer well to others due to differences in ethnicity, age, cultural context, or clinical characteristics [39,45,46]. Furthermore, methodological variability in EHR recording can introduce additional inconsistencies, even across hospitals within the same country [47]. For instance, our control group consisted of individuals receiving psychiatric treatment, reflecting a clinically realistic population in which distinctions between individuals at varying levels of suicide risk may be more subtle.

Suicide is a global public health problem, accounting for 1.5% of all deaths worldwide [3,46,48]. Public health action is urgently needed, and EHRs appear well-suited for developing suicide prevention algorithms and CDSS, as they can capture and integrate newly available clinical information in real time to help prevent deaths. The system is intended to support, not replace, clinical judgment, and its value will depend on integration into existing care pathways. In this setting, FNs are the more consequential error because missed cases may delay or prevent timely intervention, whereas FPs can be addressed through confirmatory assessments and structured follow-ups. Once a risk flag is generated, the model could trigger evidence-based responses such as safety planning-type interventions, associated with a 43% reduction in reattempts over 12 months [49], and brief contact and continuity interventions, which have been

shown to significantly reduce suicide reattempts [50]. The Catalan Suicide Risk Code provides a real-world example of this integrated pathway, linking detection with structured follow-up and continuity of care [51]. Implementation should also include ethical safeguards such as subgroup monitoring, patient transparency, and institutional governance.

Key limitations include restricted generalizability, a lack of broader psychosocial data, and a relatively small sample size reduced further by undersampling. To improve generalizability, we applied techniques such as cross-validation, separate training and testing sets, and bootstrapping. These limitations will be partially mitigated through clinical validation and large-scale implementation in the public health system, enabling evaluation in large populations. The clinical validation of these findings is a critical next step, planned as part of the second phase of this project. This will involve implementing the algorithm within the hospital information system and prospectively validating its predictions against newly collected clinical data. While we removed variables known to improve metrics, such as prior suicide attempts, with the aim of predicting nonobvious risk of suicidal behavior, known risk factors may be used in combination with the proposed algorithms to complement the detection of individuals at unknown and known risks. In summary, the use of AI-generated algorithms based on readily available and well-structured data may allow us to make progress in the identification of individuals at risk, thus increasing the chances of applying preventive measures in a timely manner.

Acknowledgments

Generative AI (ChatGPT; OpenAI) tools were used solely for (1) grammar checking and English-language correction during the preparation of the manuscript and (2) code debugging. The authors did not use generative AI for conceptualization, methodology, data analysis, interpretation of results, or writing of scientific content. All intellectual contributions and scientific reasoning remain entirely the work of the authors. The authors reviewed and verified all AI-generated code outputs before inclusion.

Funding

This research was funded by the Instituto de Salud Carlos III, Subdirección General de Evaluación y Fomento de la Investigación (ISCIII) and Fondo Europeo de Desarrollo Regional (FEDER; grant number PI21/01148), and by the Ministerio de Ciencia, Innovación y Universidades and NextGenerationEU (reference number TED2021-131979B-C31). DP thanks the support of the Spanish Ministry of Science and Innovation/ISCIII/FEDER (PI21/01148), the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement of the Generalitat de Catalunya (e-MH-PEMN-2021 SGR 01431), the CERCA Programme/Generalitat de Catalunya, the Instituto de Salud Carlos III, and the CIBER of Mental Health (CIBERSAM CB 19/09/00029). CP-G acknowledges his AI4S fellowship under the "Generación D" initiative of Red.es and the Ministerio para la Transformación Digital y de la Función Pública, for talent attraction (C005/24-ED CV1), funded by NextGenerationEU through the PRTR. XSC thanks the support of the Departament de Salut, Generalitat de Catalunya (CatSalut)–REACT NEXT GEN PT-052023-EP. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. The Department of Mental Health at the Parc Taulí University Hospital and the Institut d'Investigació i Innovació Parc Taulí (I3PT-CERCA) are the sponsors.

Data Availability

The dataset from this study is securely stored on an AQUAS (Agency for Health Quality and Assessment of Catalonia) server, with access restricted to the authors. External researchers may request access from AQUAS (padris@gencat.cat). For security reasons, all analyses must be conducted within the AQUAS server environment, and only aggregated results may be exported, under the strict supervision of AQUAS technicians. Detailed information on eligibility and access conditions is available on the AQUAS website [52].

Authors' Contributions

Conceptualization: CP-G, MF, DR, EB, CP, JG, DP
 Data curation: CP-G, XSC, MC, JFM-C, JAE, EB
 Formal analysis: CP-G, XSC, MC, JFM-C, JAE
 Funding acquisition: MF, DP
 Investigation: CP-G, MF, XSC, MC, DR, CP, JG, DP
 Methodology: CP-G, MF, XSC, MC, JFM-C, JAE, DR, EB, CP, JG, DP
 Project administration: CP-G, MF, DP
 Resources: MF, JFM-C, JAE, EB, DP
 Software: CP-G, XSC, MC, JFM-C, JAE
 Supervision: CP-G, MF, DP
 Validation: CP-G, MF, XSC, JFM-C, JAE, EB, DP
 Visualization: CP-G, XSC
 Writing – original draft: CP-G
 Writing – review and editing: CP-G, MF, CP, JG, DP

Conflicts of Interest

DP has received grants from, and served as a consultant or advisor for, Rovi, Johnson & Johnson, and Lundbeck. The other authors declare no conflicts of interest.

Multimedia Appendix 1

UpSet plot showing the distribution and overlaps of 10 key clinical variables.

[\[DOCX File \(Microsoft Word File\), 129 KB-Multimedia Appendix 1\]](#)

References

1. LIVE LIFE: an implementation guide for suicide prevention in countries. World Health Organization; 2021. URL: <https://www.who.int/publications/i/item/9789240026629> [Accessed 2025-05-01]
2. Preventing suicide: a global imperative. World Health Organization; 2014. URL: <https://iris.who.int/server/api/core/bitstreams/5bd12922-b362-458f-8025-f613da4922e7/content> [Accessed 2026-07-23]
3. Suicide worldwide in 2021: global health estimates. World Health Organization; 2025. URL: <https://iris.who.int/server/api/core/bitstreams/769d0a45-b50a-4b17-ba40-259bef44d9dd/content> [Accessed 2026-07-23]
4. Franklin JC, Ribeiro JD, Fox KR, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull.* Feb 2017;143(2):187-232. [doi: [10.1037/bul0000084](https://doi.org/10.1037/bul0000084)] [Medline: [27841450](https://pubmed.ncbi.nlm.nih.gov/27841450/)]
5. Laanani M, Imbaud C, Tuppin P, et al. Contacts with health services during the year prior to suicide death and prevalent conditions a nationwide study. *J Affect Disord.* Sep 1, 2020;274:174-182. [doi: [10.1016/j.jad.2020.05.071](https://doi.org/10.1016/j.jad.2020.05.071)] [Medline: [32469801](https://pubmed.ncbi.nlm.nih.gov/32469801/)]
6. Stene-Larsen K, Reneflot A. Contact with primary and mental health care prior to suicide: a systematic review of the literature from 2000 to 2017. *Scand J Public Health.* Feb 2019;47(1):9-17. [doi: [10.1177/1403494817746274](https://doi.org/10.1177/1403494817746274)] [Medline: [29207932](https://pubmed.ncbi.nlm.nih.gov/29207932/)]
7. Fradera M, Ouchi D, Prat O, et al. Can routine primary care records help in detecting suicide risk? A population-based case-control study in Barcelona. *Arch Suicide Res.* 2022;26(3):1395-1409. [doi: [10.1080/13811118.2021.1911894](https://doi.org/10.1080/13811118.2021.1911894)] [Medline: [33896400](https://pubmed.ncbi.nlm.nih.gov/33896400/)]
8. Bebbington PE, Minot S, Cooper C, et al. Suicidal ideation, self-harm and attempted suicide: results from the British psychiatric morbidity survey 2000. *Eur Psychiatry.* Nov 2010;25(7):427-431. [doi: [10.1016/j.eurpsy.2009.12.004](https://doi.org/10.1016/j.eurpsy.2009.12.004)] [Medline: [20627465](https://pubmed.ncbi.nlm.nih.gov/20627465/)]
9. Sveticic J, De Leo D. The hypothesis of a continuum in suicidality: a discussion on its validity and practical implications. *Ment Illn.* Jul 26, 2012;4(2):e15. [doi: [10.4081/mi.2012.e15](https://doi.org/10.4081/mi.2012.e15)] [Medline: [25478116](https://pubmed.ncbi.nlm.nih.gov/25478116/)]
10. Chan MKY, Bhatti H, Meader N, et al. Predicting suicide following self-harm: systematic review of risk factors and risk scales. *Br J Psychiatry.* Oct 2016;209(4):277-283. [doi: [10.1192/bjp.bp.115.170050](https://doi.org/10.1192/bjp.bp.115.170050)] [Medline: [27340111](https://pubmed.ncbi.nlm.nih.gov/27340111/)]
11. Large M, Kaneson M, Myles N, Myles H, Gunaratne P, Ryan C. Meta-analysis of longitudinal cohort studies of suicide risk assessment among psychiatric patients: heterogeneity in results and lack of improvement over time. *PLoS One.* 2016;11(6):e0156322. [doi: [10.1371/journal.pone.0156322](https://doi.org/10.1371/journal.pone.0156322)] [Medline: [27285387](https://pubmed.ncbi.nlm.nih.gov/27285387/)]
12. Witt K, Spittal MJ, Carter G, et al. Effectiveness of online and mobile telephone applications ('apps') for the self-management of suicidal ideation and self-harm: a systematic review and meta-analysis. *BMC Psychiatry.* Aug 15, 2017;17(1):297. [doi: [10.1186/s12888-017-1458-0](https://doi.org/10.1186/s12888-017-1458-0)] [Medline: [28810841](https://pubmed.ncbi.nlm.nih.gov/28810841/)]

13. Okolie C, Dennis M, Simon Thomas E, John A. A systematic review of interventions to prevent suicidal behaviors and reduce suicidal ideation in older people. *Int Psychogeriatr*. Nov 2017;29(11):1801-1824. [doi: [10.1017/S1041610217001430](https://doi.org/10.1017/S1041610217001430)] [Medline: [28766474](#)]
14. Milner A, Witt K, Pirkis J, et al. The effectiveness of suicide prevention delivered by GPs: a systematic review and meta-analysis. *J Affect Disord*. Mar 1, 2017;210:294-302. [doi: [10.1016/j.jad.2016.12.035](https://doi.org/10.1016/j.jad.2016.12.035)] [Medline: [28068618](#)]
15. Woodford R, Spittal MJ, Milner A, et al. Accuracy of clinician predictions of future self-harm: a systematic review and meta-analysis of predictive studies. *Suicide Life Threat Behav*. Feb 2019;49(1):23-40. [doi: [10.1111/sltb.12395](https://doi.org/10.1111/sltb.12395)] [Medline: [28972271](#)]
16. Simon GE, Matarazzo BB, Walsh CG, et al. Reconciling statistical and clinicians' predictions of suicide risk. *Psychiatr Serv*. May 1, 2021;72(5):555-562. [doi: [10.1176/appi.ps.202000214](https://doi.org/10.1176/appi.ps.202000214)] [Medline: [33691491](#)]
17. Barak-Corren Y, Castro VM, Javitt S, et al. Predicting suicidal behavior from longitudinal electronic health records. *Am J Psychiatry*. Feb 1, 2017;174(2):154-162. [doi: [10.1176/appi.ajp.2016.16010077](https://doi.org/10.1176/appi.ajp.2016.16010077)] [Medline: [27609239](#)]
18. Kessler RC, Warner CH, Ivany C, et al. Predicting suicides after psychiatric hospitalization in US Army soldiers: the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry*. Jan 2015;72(1):49-57. [doi: [10.1001/jamapsychiatry.2014.1754](https://doi.org/10.1001/jamapsychiatry.2014.1754)] [Medline: [25390793](#)]
19. Simon GE, Johnson E, Lawrence JM, et al. Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. *Am J Psychiatry*. Oct 1, 2018;175(10):951-960. [doi: [10.1176/appi.ajp.2018.17101167](https://doi.org/10.1176/appi.ajp.2018.17101167)] [Medline: [29792051](#)]
20. Poulin C, Shiner B, Thompson P, et al. Predicting the risk of suicide by analyzing the text of clinical notes. *PLoS One*. 2014;9(1):e85733. [doi: [10.1371/journal.pone.0085733](https://doi.org/10.1371/journal.pone.0085733)] [Medline: [24489669](#)]
21. Ben-Ari A, Hammond K. Text mining the EMR for modeling and predicting suicidal behavior among US veterans of the 1991 persian gulf war. Presented at: 2015 48th Hawaii International Conference on System Sciences (HICSS); Jan 5-8, 2015; Kauai, HI, USA. [doi: [10.1109/HICSS.2015.382](https://doi.org/10.1109/HICSS.2015.382)]
22. Su C, Xu Z, Pathak J, Wang F. Deep learning in mental health outcome research: a scoping review. *Transl Psychiatry*. Apr 22, 2020;10(1):116. [doi: [10.1038/s41398-020-0780-3](https://doi.org/10.1038/s41398-020-0780-3)] [Medline: [32532967](#)]
23. Walkup JT, Townsend L, Crystal S, Olfson M. A systematic review of validated methods for identifying suicide or suicidal ideation using administrative or claims data. *Pharmacoepidemiol Drug Saf*. Jan 2012;21(S1):174-182. [doi: [10.1002/pds.2335](https://doi.org/10.1002/pds.2335)] [Medline: [22262604](#)]
24. Platt R, Carnahan R. The U.S. Food and Drug Administration's Mini-sentinel Program. *Pharmacoepidemiol Drug Saf*. Jan 2012;21(S1):1-303. [doi: [10.1002/pds.3230](https://doi.org/10.1002/pds.3230)]
25. McArdle JJ, Ritschard G. *Contemporary Issues in Exploratory Data Mining in the Behavioral Sciences*. Routledge/Taylor & Francis Group; 2014. [doi: [10.4324/9780203403020](https://doi.org/10.4324/9780203403020)]
26. Walsh CG, Ribeiro JD, Franklin JC. Predicting risk of suicide attempts over time through machine learning. *Clin Psychol Sci*. May 2017;5(3):457-469. [doi: [10.1177/2167702617691560](https://doi.org/10.1177/2167702617691560)]
27. Burke TA, Ammerman BA, Jacobucci R. The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: a systematic review. *J Affect Disord*. Feb 15, 2019;245:869-884. [doi: [10.1016/j.jad.2018.11.073](https://doi.org/10.1016/j.jad.2018.11.073)] [Medline: [30699872](#)]
28. Kessler RC, Hwang I, Hoffmire CA, et al. Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans health Administration. *Int J Methods Psychiatr Res*. Sep 2017;26(3):e1575. [doi: [10.1002/mpr.1575](https://doi.org/10.1002/mpr.1575)] [Medline: [28675617](#)]
29. Zheng L, Wang O, Hao S, et al. Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records. *Transl Psychiatry*. Feb 20, 2020;10(1):72. [doi: [10.1038/s41398-020-0684-2](https://doi.org/10.1038/s41398-020-0684-2)] [Medline: [32080165](#)]
30. Edelson DP, Churpek MM, Carey KA, et al. Early warning scores with and without artificial intelligence. *JAMA Netw Open*. Oct 1, 2024;7(10):e2438986. [doi: [10.1001/jamanetworkopen.2024.38986](https://doi.org/10.1001/jamanetworkopen.2024.38986)] [Medline: [39405061](#)]
31. Winslow CJ, Edelson DP, Churpek MM, et al. The impact of a machine learning early warning score on hospital mortality: a multicenter clinical intervention trial. *Crit Care Med*. Sep 1, 2022;50(9):1339-1347. [doi: [10.1097/CCM.0000000000005492](https://doi.org/10.1097/CCM.0000000000005492)] [Medline: [35452010](#)]
32. Escobar GJ, Liu VX, Schuler A, Lawson B, Greene JD, Kipnis P. Automated identification of adults at risk for in-hospital clinical deterioration. *N Engl J Med*. Nov 12, 2020;383(20):1951-1960. [doi: [10.1056/NEJMsa2001090](https://doi.org/10.1056/NEJMsa2001090)] [Medline: [33176085](#)]
33. Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales [Article in Spanish]. *Boletín Oficial del Estado (BOE)*. URL: <https://www.boe.es/eli/es/lo/2018/12/05/3/con> [Accessed 2018-06-12]

34. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) [Article in Spanish]. European Union; 2016. URL: <https://eur-lex.europa.eu/eli/reg/2016/679/oj> [Accessed 2026-08-05]
35. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Presented at: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; Dec 4-9, 2017; Red Hook, NY, USA. [doi: [10.5555/3295222.3295230](https://doi.org/10.5555/3295222.3295230)]
36. Florek P, Zagdański A. Benchmarking state-of-the-art gradient boosting algorithms for classification. arXiv. Preprint posted online on May 26, 2023. [doi: [10.48550/arXiv.2305.17094](https://doi.org/10.48550/arXiv.2305.17094)]
37. Seyedsalehi A, Fazel S. Suicide risk assessment tools and prediction models: new evidence, methodological innovations, outdated criticisms. *BMJ Ment Health*. Mar 14, 2024;27(1):e300990. [doi: [10.1136/bmjment-2024-300990](https://doi.org/10.1136/bmjment-2024-300990)] [Medline: [38485246](https://pubmed.ncbi.nlm.nih.gov/38485246/)]
38. Ehtemam H, Sadeghi Esfahlani S, Sanaei A, et al. Role of machine learning algorithms in suicide risk prediction: a systematic review-meta analysis of clinical studies. *BMC Med Inform Decis Mak*. May 27, 2024;24(1):138. [doi: [10.1186/s12911-024-02524-0](https://doi.org/10.1186/s12911-024-02524-0)] [Medline: [38802823](https://pubmed.ncbi.nlm.nih.gov/38802823/)]
39. Pigoni A, Delvecchio G, Turtulici N, et al. Machine learning and the prediction of suicide in psychiatric populations: a systematic review. *Transl Psychiatry*. Mar 9, 2024;14(1):140. [doi: [10.1038/s41398-024-02852-9](https://doi.org/10.1038/s41398-024-02852-9)] [Medline: [38461283](https://pubmed.ncbi.nlm.nih.gov/38461283/)]
40. Kirtley OJ, van Mens K, Hoogendoorn M, Kapur N, de Beurs D. Translating promise into practice: a review of machine learning in suicide research and prevention. *Lancet Psychiatry*. Mar 2022;9(3):243-252. [doi: [10.1016/S2215-0366\(21\)00254-6](https://doi.org/10.1016/S2215-0366(21)00254-6)] [Medline: [35183281](https://pubmed.ncbi.nlm.nih.gov/35183281/)]
41. Lagerberg T, Gustafsson TT, Molero Y, et al. Oral glucocorticoids and risk of psychiatric and suicidal behaviour outcomes: population-based cohort study. *Br J Psychiatry*. Jul 2026;229(1):43-50. [doi: [10.1192/bjp.2025.128](https://doi.org/10.1192/bjp.2025.128)] [Medline: [40556309](https://pubmed.ncbi.nlm.nih.gov/40556309/)]
42. Lin HT, Lai CH, Perng HJ, et al. Insomnia as an independent predictor of suicide attempts: a nationwide population-based retrospective cohort study. *BMC Psychiatry*. May 2, 2018;18(1):117. [doi: [10.1186/s12888-018-1702-2](https://doi.org/10.1186/s12888-018-1702-2)] [Medline: [29716570](https://pubmed.ncbi.nlm.nih.gov/29716570/)]
43. Piera-Jiménez J, Carot-Sans G, Ramiro-Pareta M, et al. A 25-year retrospective of health IT infrastructure building: the example of the catalonia region. *J Med Internet Res*. Nov 18, 2024;26:e58933. [doi: [10.2196/58933](https://doi.org/10.2196/58933)] [Medline: [39556831](https://pubmed.ncbi.nlm.nih.gov/39556831/)]
44. Peña Gómez C, Fradera M, Caravaca M, Roche D, Giraldo J, Palao D. Machine learning model development to retrospectively predict suicide attempts in the Millenium Cohort Study sample. *Eur Psychiatr*. Apr 2025;68(S1):S120. [doi: [10.1192/j.eurpsy.2025.335](https://doi.org/10.1192/j.eurpsy.2025.335)]
45. Haroz EE, Rebman P, Goklish N, et al. Performance of machine learning suicide risk models in an American Indian population. *JAMA Netw Open*. Oct 1, 2024;7(10):e2439269. [doi: [10.1001/jamanetworkopen.2024.39269](https://doi.org/10.1001/jamanetworkopen.2024.39269)] [Medline: [39401036](https://pubmed.ncbi.nlm.nih.gov/39401036/)]
46. Martínez-Alés G, Jiang T, Keyes KM, Gradus JL. The recent rise of suicide mortality in the United States. *Annu Rev Public Health*. Apr 5, 2021;43:99-116. [doi: [10.1146/annurev-publhealth-051920-123206](https://doi.org/10.1146/annurev-publhealth-051920-123206)] [Medline: [34705474](https://pubmed.ncbi.nlm.nih.gov/34705474/)]
47. Singh H, Mhasawade V, Chunara R. Generalizability challenges of mortality risk prediction models: a retrospective analysis on a multi-center database. *PLOS Digit Health*. Apr 2022;1(4):e0000023. [doi: [10.1371/journal.pdig.0000023](https://doi.org/10.1371/journal.pdig.0000023)] [Medline: [36812510](https://pubmed.ncbi.nlm.nih.gov/36812510/)]
48. Vos T, Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*. Oct 2020;396(10258):1204-1222. [doi: [10.1016/S0140-6736\(20\)30925-9](https://doi.org/10.1016/S0140-6736(20)30925-9)]
49. Nuij C, van Ballegooijen W, de Beurs D, et al. Safety planning-type interventions for suicide prevention: meta-analysis. *Br J Psychiatry*. Aug 2021;219(2):419-426. [doi: [10.1192/bjp.2021.50](https://doi.org/10.1192/bjp.2021.50)] [Medline: [35048835](https://pubmed.ncbi.nlm.nih.gov/35048835/)]
50. Homan S, Marciniak MA, Michel S, et al. Effectiveness of brief interventions and contacts after suicide attempt: a systematic review and meta-analysis. *EClinicalMedicine*. Mar 2026;93:103824. [doi: [10.1016/j.eclim.2026.103824](https://doi.org/10.1016/j.eclim.2026.103824)] [Medline: [41859679](https://pubmed.ncbi.nlm.nih.gov/41859679/)]
51. Pérez V, Elices M, Prat B, et al. The Catalonia Suicide Risk Code: a secondary prevention program for individuals at risk of suicide. *J Affect Disord*. May 1, 2020;268:201-205. [doi: [10.1016/j.jad.2020.03.009](https://doi.org/10.1016/j.jad.2020.03.009)] [Medline: [32174478](https://pubmed.ncbi.nlm.nih.gov/32174478/)]
52. Agència de Qualitat i Avaluació Sanitàries de Catalunya (AQuAS) [Article in Catalan]. URL: <https://aquas.gencat.cat/ca/inici> [Accessed 2026-07-23]

Abbreviations

AdaBoost: adaptive boosting

ADASYN: Adaptive Synthetic Sampling

AQUAS: Agency for Health Quality and Assessment of Catalonia

CDSS: clinical decision support system

DL: deep learning

EHR: electronic health record

EWS: early warning system

FN: false-negative

FP: false-positive

ML: machine learning

ROC-AUC: receiver operating characteristic-area under the curve

RWD: real-world data

SHAP: Shapley additive explanations

SMOTE: Synthetic Minority Oversampling Technique

TN: true-negative

TP: true-positive

XGBoost: extreme gradient boosting

Edited by Yanshan Wang; peer-reviewed by Eric Pelletier, Kunal Khashu; submitted 30.Sep.2025; final revised version received 01.May.2026; accepted 15.Jun.2026; published 26.Aug.2026

Please cite as:

Peña-Gómez C, Fradera M, Sánchez Corrales X, Caravaca-Rodriguez M, Martínez-Cerdá JF, Escofet JA, Roche D, Barbería E, Pontes C, Giraldo J, Palao D

AI Models for Suicide Risk Prediction in Adult Patients Receiving Mental Health Care Using Real-World Data: Retrospective Population-Based Study

JMIR AI 2026;5:e85085

URL: <https://ai.jmir.org/2026/1/e85085>

doi: [10.2196/85085](https://doi.org/10.2196/85085)

© Cleofé Peña-Gómez, Marc Fradera, Xavier Sánchez Corrales, Marc Caravaca-Rodriguez, Juan-Francisco Martínez-Cerdá, Joan Albert Escofet, David Roche, Eneko Barbería, Caridad Pontes, Jesús Giraldo, Diego Palao. Originally published in JMIR AI (<https://ai.jmir.org>), 26.Aug.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 AI, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.ai.jmir.org/>, as well as this copyright and license information must be included.