GOIANA JOURNAL OF MEDICINE

ISSN 0034-9585
SCIENTIFIC ARTICLE - REVISION

ARTIFICIAL INTELLIGENCE AND SUICIDE RISK: PREDICTIVE LIMITS AND COMPLEMENTARY CLINICAL POTENTIAL

KALEBE DIAS DA SILVA[1], ANA LUIZA NAUJORKS ZIMMER[1], BIANCA DE SOUZA FERREIRA[2], JOÃO HENRIQUE PAZ DA SILVA RIBEIRO[1], FABIANO INÁCIO DE SOUZA[3]

1. Graduando(a) em Medicina Universidade Federal de Goiás, Goiânia - GO, Brazil.
2. Graduanda em Medicina pelo Centro Universitário Goyazes - Unigoyazes, Trindade - GO, Brazil.
3. Doutor em Ortopedia, Traumatologia e Reabilitação pelaUniversidade Federal de Goiás, Goiânia - GO, Brazil.

ABSTRACT

This study is an integrative literature review designed to synthesize evidence on the accuracy and performance of Artificial Intelligence and Machine Learning models in predicting suicide risk. The methodological design strictly followed the PRISMA 2020 guidelines. Searches were conducted in the PubMed/MEDLINE, Scopus, Web of Science, and Embase databases. The audit of methodological quality and risk of bias used the PROBAST+AI tool. The final sample included 53 primary studies with consistent data for performance evaluation. The results demonstrated that the pooled value of the models reached a global Area Under the Curve (AUC-ROC) of 0.86, indicating good overall predictive capacity. Despite this, a paradox was observed between statistical prediction and clinical utility. The Positive Predictive Value (PPV) scored around 0.01 in most tests, representing high false-positive rates. Furthermore, the application of the PROBAST tool revealed that approximately 93% of the evaluated models are at a high risk of bias. In conclusion, Artificial Intelligence should be understood as a complementary tool to clinical assessment, contributing to early identification. However, limitations related to the clinical utility of some models, the absence of external validation, and the predominance of retrospective studies still restrict its broad application in clinical practice.

Keywords: Artificial intelligence, Machine learning, Risk assessment, Suicide.

INTRODUCTION

Suicide remains a challenging epidemiological issue today, particularly in Brazil. According to data published by the Brazilian Psychiatric Association (ABP), more than 144,000 deaths were recorded between 2013 and 2023, with a rising national mortality rate.[1] In 2013, the rate was 5.27 deaths per 100,000 inhabitants, increasing to 7.50 in 2023, with a continued upward trend. Furthermore, the young age profile of this population is noteworthy, particularly individuals aged 20 to 49, who account for 59% of the victims.[1]

Traditionally, the prediction of suicidal behavior in medical practice has relied on cross-sectional clinical assessment and the identification of established risk factors. However, the multifactorial,

dynamic, and often acute nature of suicidal ideation imposes limitations on the predictive accuracy of conventional methods. In this context, integrating technological innovations into the prediction, screening, and prevention of this outcome has become a necessity. With the rise of Artificial Intelligence (AI) and Large Language Models (LLMs), various algorithmic tools, particularly those based on Machine Learning, have been developed to enable early identification of the risk of suicidal ideation, suicide attempts, and completed suicide.

However, given the recent introduction of these tools into the healthcare ecosystem, a rigorous evaluation of their performance metrics is required. It is imperative to critically analyze the sensitivity, specificity, and Positive Predictive Value (PPV) of these models, as well as to investigate the risk of bias inherent in the algorithms. Above all, it is necessary to question the external validity and actual clinical applicability of these technologies, determining whether they truly represent a tangible and safe benefit for medical decision-making and patient care.

OBJECTIVES

The primary objective of this study is to synthesize scientific evidence on the accuracy and performance of Artificial Intelligence and Machine Learning models in predicting suicide risk. As secondary objectives, this integrative review aims to: (I) evaluate the clinical applicability of these tools compared to traditional psychiatric screening methods; and (II) critically analyze the limitations and main methodological biases present in the current literature on the subject.

METHODOLOGY

This study is an integrative literature review conducted with the aim of synthesizing evidence and evaluating the accuracy of Artificial Intelligence (AI) models in predicting suicide risk.[2] The methodological design strictly followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines, ensuring the transparency and reproducibility of the research.[4] As a secondary review study based on publicly available literature data, the study was exempt from direct submission to the Research Ethics Committee (CEP), in full compliance with the guidelines of National Health Council (CONEP) Resolution 466/2012.[3] The formulation of the research question and the definition of eligibility criteria were based on the PICOTS acronym (Population, Intervention/Index Model, Comparator, Outcomes, Timing, and Setting).[4]

The population included individuals of all age groups; the intervention encompassed AI algorithms, such as Machine Learning; the comparator consisted of traditional clinical risk assessment; and the primary outcomes included suicidal ideation, suicide attempts, and completed suicide.[5] The search strategy was extensively applied across the following electronic databases: PubMed/MEDLINE, Scopus, Web of Science, and Embase.[5] Controlled vocabularies, including Medical Subject Headings (MeSH), were used in combination with free-text terms.[6] The search equation combined technological and clinical constructs using the Boolean operators "AND" and "OR", resulting in search strings that included "Artificial Intelligence", "Machine Learning", "Suicide", and "Risk Assessment".[6] No restrictions were applied regarding the year of publication, and studies written predominantly in English were included. Article selection was conducted in two main stages: initial screening of titles and abstracts, followed by full-text review of the selected manuscripts.

Inclusion or exclusion decisions were made by two independent human reviewers, with disagreements resolved by a third senior reviewer. The exclusion criteria included non-original or incomplete

publications, methodologies that did not employ Machine Learning tools, outcomes not focused on suicide attempts or lethality, limited data (self-reported data or aggregated predictors), and overlapping samples or the absence of clearly defined metrics.

Data extraction and the assessment of methodological quality and risk of bias in the selected studies were performed using the PROBAST+AI (Prediction model Risk Of Bias Assessment Tool for Artificial Intelligence) tool.[5] This instrument enabled an in-depth analysis across four critical domains: participant selection, predictors, outcomes, and statistical analysis, focusing on the detection of vulnerabilities inherent in AI.[4] The results were organized into a synthesis matrix through the extraction of predictive metrics, such as the Area Under the ROC Curve (AUC), sensitivity, specificity, and positive predictive value. Subsequently, the extracted data were analyzed descriptively and categorized qualitatively to develop an integrative synthesis of the existing knowledge on the subject.[7]

RESULTS

In this integrative review, the initial search of electronic databases (PubMed, Scopus, Web of Science, and Embase) resulted in the identification of 16,217 records.[2] After duplicate removal, 15,550 titles and abstracts were screened, of which 14,888 were excluded for not meeting the predefined eligibility criteria.[2] Subsequently, 662 full-text articles were retrieved for assessment, of which 534 were accessed and evaluated.[2] The review of these publications led to the exclusion of studies that did not use Machine Learning models, did not assess suicidal ideation, suicide attempts, or completed suicide, or did not provide comparisons with traditional clinical methods of suicide risk prediction. At the end of this process, 163 studies comprised the qualitative synthesis sample of this integrative review.[2]

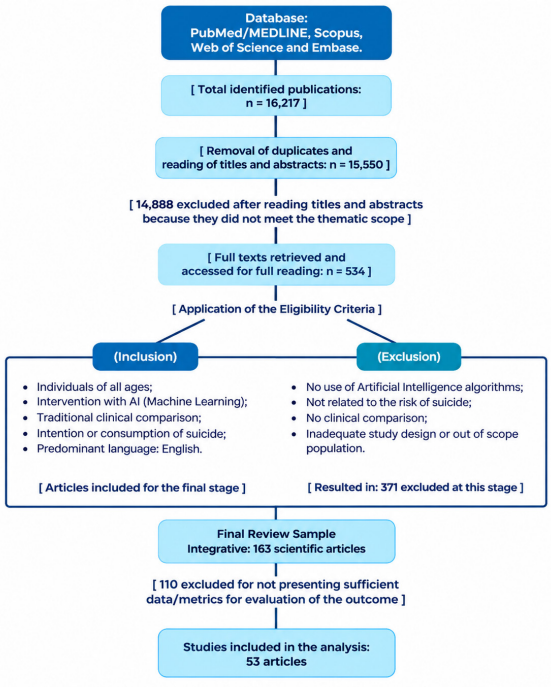


Figure 1: Flowchart of study selection based on PRISMA 2020..
Source: Prepared by the authors using data from PubMed, Scopus, Web of Science, and Embase.

The final sample therefore comprised 53 primary studies that presented consistent data for performance evaluation. [3] The relevance of the selected articles lies in their attempt to improve and understand suicide risk prediction through Artificial Intelligence algorithms, such as Machine Learning. However, the current literature was found to exhibit considerable methodological heterogeneity. [4] Thus, given the lack of global standardization in the extraction of demographic and geographic variables, this review focused on the intrinsic quality and accuracy metrics of the models developed.

The application of the PROBAST+AI tool revealed that only 8% of the developed AI models had undergone external validation and presented a low risk of bias in their validations.[4] It also revealed that approximately 93% of the models developed and evaluated were at high risk of bias (High Risk of Bias).[5] This risk was primarily attributable to shortcomings in the analysis domain (Domain 4). The downgrading resulted from inadequate assessment of predictive performance in 92% of the studies, combined with insufficient sample size (77%) and inappropriate handling of missing data (66%).[4]

The discriminative ability of Machine Learning algorithms demonstrated positive results, despite variations in some of the metrics analyzed in Table 1. The overall Area Under the Receiver Operating Characteristic Curve (AUC-ROC) ranged from 0.69 to 0.93.[3] The pooled value of the models reached an overall AUC of 0.86, indicating good overall predictive performance.[5] When stratified by outcome, the AUC showed consistent results: 0.88 for suicidal ideation, 0.87 for suicide attempts, and 0.84 for completed suicide.[5]

Table 1: Flowchart of study selection based on PRISMA 2020.

Accuracy Metric	Variation (Minimum - Maximum)	Grouped Average (Pooled)
Area Under the Curve (AUC-ROC) – Global	0.69 – 0.93	0.86
AUC: Suicide Ideation		0.88
AUC: Suicide Attempt		0.87
AUC: Consumed Suicide		0.84
Sensitivity	45% – 82%	0.66
Specificity	91% – 95%	0.87
Positive Likelihood Ratio (LR+)	6.5 – 9.9	
Negative Likelihood Ratio (LR-)	0.2 – 0.6	
Positive Predictive Value (PPV)	6% – 17%	

Source: Prepared by the authors using data from PROBAST+AI.

DISCUSSION

The The results demonstrated that Artificial Intelligence, through Machine Learning, shows promising performance in predicting suicide risk, as it is capable of predicting future suicidal behaviours and facilitating early support and intervention.[2] In this context, the use of Artificial Intelligence for this purpose achieved accuracy levels exceeding 90% and may outperform traditional prediction models, as these are based on subjective assessments, which may vary depending on the professional's previous experience.[2] However, the authors emphasised that this technology should not replace traditional methods but rather serve as a complementary tool to improve the identification of suicidal behaviours in clinical practice.[2]

From this perspective, despite the considerable overall accuracy rates reported in the selected studies, the literature indicates a paradox between statistical prediction and the clinical utility of these tools. As demonstrated through simulations, although many AI models achieve an overall accuracy above 0.8, their practical utility in predicting suicide mortality was found to be almost negligible.[8] In this regard, the Positive Predictive Value (PPV) of these technologies was approximately 0.01 in most tests, indicating high false-positive rates. This phenomenon reflects a consequence of the low suicide mortality rate in the general population.

In this context, the predominant methodological approach used to develop AI algorithms was supervised learning, with an emphasis on the application of Random Forests and Support Vector Machines (SVMs).[6] Thus, Electronic Medical Records (EMRs) were identified as the primary data source for these systems.[6] Furthermore, the use of Natural Language Processing (NLP) was observed, a tool capable of detecting suicidal tendencies and behaviours through the analysis of texts, comments and user interactions on social media.[2] Some studies also employed NLP-based Textual Emotion Detection techniques, which analyse users' texts and identify individuals at risk of suicide.[2] The Reddit platform, for example, allows users to post anonymously and provides a virtual environment conducive to expressing their difficulties and emotional problems. Consequently, AI can facilitate the early identification of signs of suicidal ideation through subtle changes in language use, enabling assistance to be provided to these individuals when necessary.

It is important to highlight that a study conducted with chatbots that establish personalised relationships with users detected 44% of suicidal thoughts and, in these cases, 88% of these platforms referred users to crisis support services.[8] By providing a welcoming and non-judgemental environment, these chats offer opportunities for individuals to express their emotions freely. Thus, through interactions with their users, AI can identify mental health problems, as well as mood disorders and suicidal ideation. In such cases, after detecting these changes, it serves as an essential tool for guiding users towards appropriate professional care.

Although Machine Learning has limitations in predicting exact suicide mortality, it demonstrates greater utility in identifying suicide attempts, revealing an important temporal dynamic in the progression of risk.[6] From this perspective, studies illustrate that the relevance of predictive factors changes as a suicide attempt becomes more imminent.[7] Through a broad analysis covering months or years before the event, factors such as a history of hospital utilisation and psychiatric diagnoses (such as depression and schizophrenia) are determining factors. However, over shorter time frames, substance dependence diagnoses acquire greater predictive weight.[7] Thus, although the algorithm identifies individuals at long-term risk of suicide, it lacks the sensitivity required to determine imminent suicide risk. In this

regard, due to this limitation, healthcare professionals cannot base suicide risk prediction solely on AI. Therefore, the technology does not replace the knowledge and experience of healthcare professionals in identifying signs or symptoms during medical practice.

Through the analysis of demographic subgroups, the studies demonstrated that adolescence is considered a period of high vulnerability, requiring the use of prediction tools informed by the analysis of behaviours specific to this group.[6] In this regard, the severity of depressive symptoms emerges as the strongest predictor of suicidal behaviours.[6] Furthermore, other factors, such as drug abuse and high levels of academic stress, constitute important predictive variables for training the algorithms.[6] Notably, the literature demonstrates that prolonged experiences of bullying, both physical and cyberbullying, seriously affect adolescents' mental health, increasing psychological distress and isolation, factors that contribute to the development of future suicidal ideation.[6] However, it is essential to emphasise that the effectiveness of AI in early identification does not eliminate the active role of society. In this context, families and school communities play an indispensable role in supporting young people, contributing significantly to mediating and combating stressors such as bullying, with the aim of promoting a supportive environment capable of preventing the progression of suicidal behaviour.

In addition to clinical factors, the accuracy of Artificial Intelligence in predicting suicide among adolescents is related to its ability to adapt to users' individual and sociocultural differences. In this regard, the application of these tools revealed that girls have detection rates for suicidal ideation and attempts that are twice as high, whereas boys exhibit greater lethality due to their choice of more aggressive methods.[6] Furthermore, cultural diversity affects the generalisability of predictive models, since the relevance of factors such as family relationships may differ according to the social context. Therefore, AI models are susceptible to errors if they are not adjusted to account for diverse social settings and the specific family dynamics of each individual.

Despite the excellent predictive potential of Artificial Intelligence in suicide prediction, the experimental use of this technology in medical practice has methodological limitations in the current literature. Most studies are based on cross-sectional and retrospective designs, making their conclusions susceptible to recall and selection bias. Furthermore, it is important to highlight a limitation related to coding in electronic medical records, as only a small proportion of suicide attempts receive the appropriate International Classification of Diseases (ICD) code, making it impossible to train algorithms based solely on this information.[6] In addition, the lack of external validation also limits the use of these tools, as few studies have tested their models outside the original datasets.

CONCLUSION

Considering the results of the present integrative review, it can be concluded that late complications associated with the use of hyaluronic acid dermal fillers in facial aesthetic procedures are multifactorial events influenced by injection technique, biomaterial characteristics, individual inflammatory response, and the patient's systemic conditions. The main adverse events described include intermittent delayed edema, inflammatory nodules, granulomas, delayed hypersensitivity reactions, and conditions associated with biofilm formation.

Therefore, it is recommended that professionals responsible for injectable aesthetic procedures adopt individualized clinical assessment, in-depth anatomical knowledge, careful product selection, and protocols for the prevention and early management of complications. In addition, further prospective studies with larger numbers of patients and stratification of risk factors are needed in order to improve

safety protocols and strengthen evidence-based clinical practice.

REFERENCES

1. Lee SY, Youn CS, Park KY. Midface volumization and skin quality enhancement with a hyaluronic acid-polynucleotide hybrid filler: 24-week prospective open-label study. J Dermatolog Treat. 2026;37(1):2621537. doi:10.1080/09546634.2026.2621537.

2. Urdiales-Gálvez F, Delgado NE, Figueiredo V, Lajo-Plaza JV, Mira M, Ortíz-Martí F, Del Rio-Reyes R, Romero-Álvarez N, Ruiz del Cueto S, Segurado MA, Villanueva Rebenaque C. Preventing the complications associated with the use of dermal fillers in facial aesthetic procedures: an expert group consensus report. Aesthetic Plast Surg. 2017;41(3):667-77. doi:10.1007/s00266-017-0798-y.

3. Piton LMM, Freitas MFL, Oliveira CMS, Silva CF. Principais complicações associadas ao preenchimento facial com ácido hialurônico: uma revisão de literatura. Rev Ibero-Am Humanid Cienc Educ. 2025;11(11):5681-91. doi:10.51891/rease.v11i11.22348.

4. Ciaramicolo NO, Bisson GB, Piedade EFDS, Osny FJ. Late facial edema after lip filling with hyaluronic acid: possible association with the use of Ozempic. J Craniofac Surg. 2024;35(7):2110-2. doi:10.1097/SCS.0000000000010588.

5. Qiu H, Xu J, Wu X, Wu Y, Chen X, Zhong P, Wang F, Lin Q, Li F, Xu T. Clinical outcomes of intense pulsed light combined with non-crosslinked sodium hyaluronate, tranexamic acid and vitamin C mesotherapy for facial photoaging: a retrospective study. J Dermatolog Treat. 2026;37(1):2609454. doi:10.1080/09546634.2025.2609454.

6. Shturman I, Garzón H, Coimbra D, Villafranca J, Zimbres S, Pyke MLA, Van Der Graaff I, Dabronzo ML. Minimally invasive aesthetic treatments for the corrective patient archetype: transforming lives. Aesthetic Plast Surg. 2026;50(11):4461-75. doi:10.1007/s00266-026-05747-w.

7. Bettiol NB, de Carvalho FA, Siessere S, Romero GDA, de Menezes M, Martini CCJ, Mazzi-Chaves JF, Magri LV, Regalo SCH, Palinkas M. Assessment of orofacial function after mandibular angle harmonization with hyaluronic acid: an observational longitudinal study. Dent J (Basel). 2026;14(4):241. doi:10.3390/dj14040241.

8. Xue J, Zhang B, Zhao Y, Zhang Q, Zheng C, Jiang J, Li H, Liu N, Li Z, Fu W, Peng Y, Logan J, Zhang J, Xiang X. Evaluation of the Current State of Chatbots for Digital Health: Scoping Review. J Med Internet Res [Internet]. 2023 [Citado em 3 Jun 2026]; 25: e47217. Disponível em: https://doi.org/10.2196/47217.

MAILING ADDRESS

KALEBE DIAS DA SILVA
Rua Curitiba, quadra 29 lote 3, Jardim Belo Horizonte, Aparecida de Goiânia/GO, Brazil.
E-mail: kalebesilva@discente.ufg.br

EDITORIAL AND REVIEW

Chief editors:
Waldemar Naves do Amaral - http://lattes.cnpq.br/4092560599116579 - https://orcid.org/0000-0002-0824-1138
Nílzio Antônio da Silva - http://lattes.cnpq.br/1780564621664455 - https://orcid.org/0000-0002-6133-0498

Authors:
Kalebe Dias da Silva - http://lattes.cnpq.br/6143323506263868 - https://orcid.org/0009-0009-2327-3494

Ana Luiza Naujorks Zimmer - http://lattes.cnpq.br/8076289651834199 - https://orcid.org/0000-0001-8055-9821

Bianca de Souza Ferreira - http://lattes.cnpq.br/9028110575514969 - https://orcid.org/0009-0004-3457-9724

João Henrique Paz da Silva Ribeiro - http://lattes.cnpq.br/2383483197313148 - https://orcid.org/0009-0001-4384-8584

Fabiano Inácio de Souza - http://lattes.cnpq.br/8305270225089632 - https://orcid.org/0000-0001-5862-4674

Library Review: Izabella Goulart
Spell Check: Dario Alvares
Translation: Soledad Montalbetti
Received: 24/08/26. Accepted: 27/08/26. Published in: 09/11/2026.