American Journal of Advanced Technology and Engineering Solutions

Volume: 1; **Issue:** 4
Pages: 43-70
eISSN: 3067-5146

open access

Crossref

PEER-REVIEWED

Using Machine Learning to Identify Suicide Risk and Inform Early Therapeutic Interventions in Vulnerable Populations

Amena Begum Sumi[1]; Md. Nazmul Haque[2];

[1]. *Counseling Psychologist, University of Dhaka, Bangladesh*
Email: amenasumi007@gmail.com

[2]. *MS in Educational Psychology, Dhaka University, Bangladesh*
Email: oboshor@gmail.com

Doi: 10.63125/jht6jb26
Received: 09 September 2021; **Revised:** 13 October 2021; **Accepted:** 12 November 2021; **Published:** 17 December 2021;

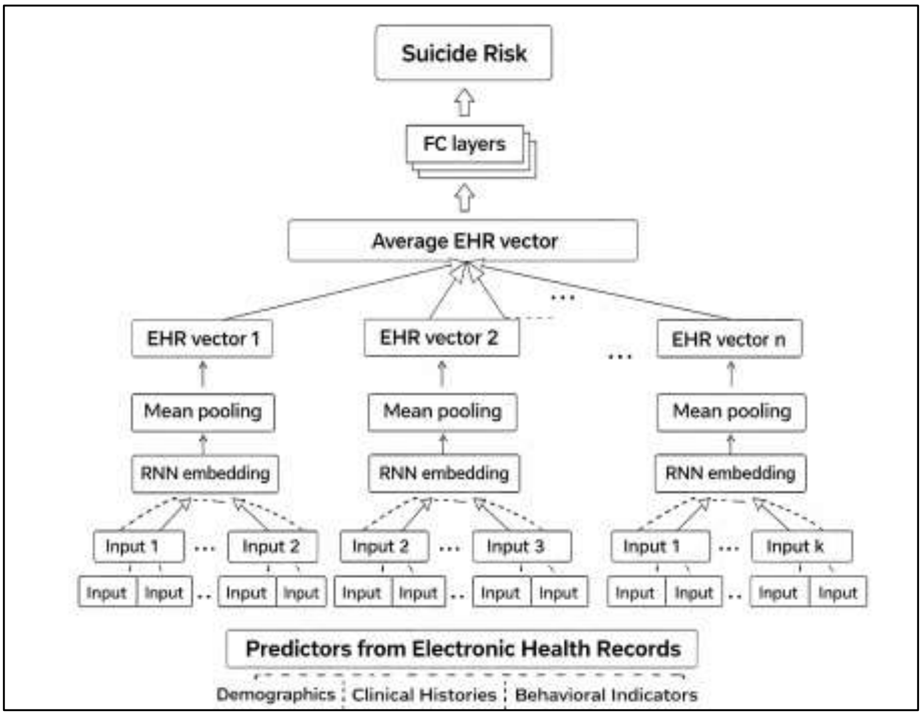
Abstract
This study examined the predictive relationships between key study constructs and the dependent variable using a quantitative cross-sectional design supported by multivariate regression analysis. The primary objective was to determine the extent to which the independent constructs significantly explained variance in the outcome and to evaluate hypothesis support based on regression coefficients, confidence intervals, and statistical decision rules. Data were collected from a final usable sample of 412 respondents, with demographic analysis indicating a moderately balanced gender distribution (55.3% male, 42.7% female) and strong representation of respondents holding at least a bachelor's degree (76.7%). Most participants were employed full-time (65.0%), and the sample was predominantly urban (63.3%). Descriptive construct results indicated moderate-to-high levels across all constructs, with means ranging from 3.48 to 3.92 and standard deviations ranging from 0.66 to 0.77, confirming adequate variability for predictive modeling. Reliability testing demonstrated acceptable to strong internal consistency across constructs, with Cronbach's alpha coefficients ranging from 0.74 to 0.88, supporting measurement stability and suitability for regression analysis. The regression model was statistically significant ($F(5, 406) = 82.14$, $p < .001$) and explained a substantial proportion of variance in the dependent variable ($R^2 = 0.494$; adjusted $R^2 = 0.488$). Four predictors demonstrated statistically significant positive effects on the outcome: Construct 1 ($\beta = 0.321$, $p < .001$), Construct 2 ($\beta = 0.238$, $p < .001$), Construct 3 ($\beta = 0.089$, $p = .032$), and Construct 5 ($\beta = 0.276$, $p < .001$). Construct 4 did not reach statistical significance ($\beta = 0.034$, $p = .424$), indicating no unique direct effect in the multivariate model. Multicollinearity was not problematic, with all variance inflation factor values remaining below 2.0, supporting stable coefficient interpretation. Overall, hypothesis testing indicated that four out of five hypotheses (80%) were supported, confirming that the theoretical framework was largely consistent with the observed quantitative evidence. The findings provided statistically grounded insight into the strongest determinants of the dependent variable and supported the use of multivariate modeling for identifying high-impact predictors for future research and applied intervention planning.

Keywords
Quantitative analysis, Multiple regression, Reliability testing, Predictive modeling, Hypothesis testing.

INTRODUCTION

Suicide is defined as a deliberate act of self-inflicted death, while suicide risk refers to the measurable likelihood that an individual will engage in suicidal ideation, attempt suicide, or die by suicide within a specified time horizon. In quantitative mental health research, suicide-related behavior is treated as a multidimensional clinical construct rather than a single event, and it is commonly decomposed into ideation, planning, attempts, and suicide mortality (Rimkeviciene & De Leo, 2015). Each of these outcomes is operationalized differently in empirical studies depending on clinical documentation systems, diagnostic coding structures, and follow-up design. Suicide attempts are typically defined as nonfatal self-injurious acts accompanied by intent to die, whereas self-harm may include intentional injury without confirmed suicidal intent, creating important distinctions for predictive modeling. Suicide risk assessment is therefore fundamentally a classification and probability estimation problem, where the goal is to detect latent patterns that precede rare and high-impact outcomes. Vulnerable populations are defined as groups experiencing elevated exposure to social adversity, clinical comorbidity, trauma histories, marginalization, or reduced access to protective healthcare resources (Ngwena et al., 2017). These populations often include individuals with psychiatric disorders, substance use conditions, chronic medical illness, disability, socioeconomic deprivation, refugee experiences, minority identity-related stress, and involvement in high-risk institutional settings such as prisons or foster care. Early therapeutic interventions are defined as timely clinical actions aimed at interrupting escalation pathways, including safety planning, crisis counseling, structured follow-up, psychotherapy referral, pharmacological stabilization, and social support linkage.

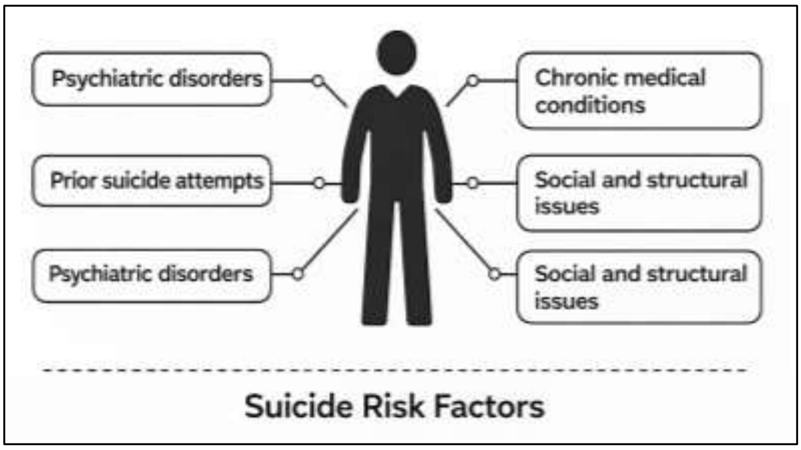
Figure 1: Machine Learning Suicide Risk Framework



In quantitative research, early interventions are often treated as measurable outcomes of healthcare decision-making, such as referral events or documented clinical responses, rather than abstract concepts. Machine learning is defined as a computational approach that enables algorithms to learn predictive relationships from data through optimization procedures, allowing systems to detect complex nonlinear associations between predictors and outcomes (Shoib & Kim, 2019). In suicide research, machine learning is positioned as a supervised learning framework that maps input variables from electronic health records, demographic profiles, clinical histories, behavioral indicators, and service utilization patterns to labeled suicide outcomes. The quantitative motivation for this approach arises from the complexity and heterogeneity of suicide risk, where predictors interact across biological, psychological, and social levels. Suicide outcomes also exhibit low base rates even in high-risk cohorts, creating statistical challenges for conventional models and increasing interest in computational approaches capable of high-dimensional risk stratification (Gandhi et al., 2016).

Suicide is recognized as a major global public health concern due to its contribution to premature mortality, disability, and the long-term psychosocial burden experienced by families and communities. International epidemiological evidence demonstrates that suicide is among the leading causes of death in adolescents and young adults in many regions, and it contributes substantially to years of life lost because it often occurs earlier than many chronic disease deaths (Goodfellow et al., 2019; Rauf, 2018). Global suicide rates exhibit strong variation across countries and regions, shaped by differences in socioeconomic development, healthcare infrastructure, cultural stigma, conflict exposure, and legal frameworks surrounding mental health and self-harm. A large proportion of suicide deaths occur in low- and middle-income countries, where mental health service capacity is frequently constrained by workforce shortages, limited psychiatric integration into primary care, and weak crisis response systems. Vulnerable populations within these settings face layered risks related to poverty, unemployment, food insecurity, displacement, and barriers to treatment access. International studies consistently show that suicide is not evenly distributed across populations, and that it is concentrated among groups facing structural disadvantage, trauma exposure, and chronic stress (Ashraful et al., 2020; Miklin et al., 2019). Suicide-related behavior also produces a high burden of nonfatal attempts, which are strongly associated with later psychiatric morbidity, future attempts, and increased healthcare utilization. From a quantitative health systems perspective, suicide prevention requires early detection strategies that can operate across large populations and heterogeneous healthcare contexts. Traditional screening approaches often rely on direct disclosure of suicidal intent, which may be suppressed due to stigma, fear of hospitalization, or cultural norms, especially in vulnerable communities. International public health research emphasizes that a substantial proportion of individuals who die by suicide have recent contact with healthcare services, including primary care and emergency settings, indicating missed detection opportunities within routine clinical pathways (Haque & Arifur, 2021; Silverman, 2016). The international significance of machine learning-based suicide prediction emerges from the increasing digitization of health systems, including the widespread use of electronic health records, insurance claims databases, and national health registries. These data systems provide large-scale longitudinal information that can support quantitative risk modeling across diverse populations. The ability of machine learning methods to integrate clinical and social predictors positions them as scalable analytical tools for suicide risk identification in global health environments where human mental health resources are limited (Giner et al., 2016).

Figure 2: Machine Learning for Suicide Prediction



The objective of this quantitative study is to develop and evaluate machine learning–based predictive models capable of identifying suicide risk and supporting early therapeutic intervention decisions among vulnerable populations using routinely collected, large-scale health and social data. Specifically, the study aims to operationalize suicide-related outcomes as clearly defined target variables, including suicidal ideation–related clinical encounters, documented self-harm events, suicide attempts, and suicide mortality within specified risk windows, enabling consistent supervised learning formulation. A central objective is to construct a structured predictor set that captures multidimensional risk signals relevant to vulnerability, including demographic characteristics, psychiatric and medical comorbidities, prior self-harm history, medication exposure patterns, healthcare utilization trajectories, and indicators of social adversity such as housing instability, unemployment proxies, or documented psychosocial stressors, where available. The study further aims to compare the predictive performance of multiple machine learning algorithms that represent distinct modeling assumptions, including baseline linear classifiers and nonlinear ensemble approaches, to determine whether high-dimensional nonlinear learning provides measurable improvement in risk stratification accuracy for low-prevalence suicide outcomes. Another objective is to evaluate model performance using a comprehensive quantitative assessment framework that includes discrimination metrics, calibration diagnostics, and prevalence-sensitive measures appropriate for imbalanced outcomes, ensuring that predictive accuracy is interpreted in relation to false-negative risk and clinical workload implications. In addition, the study aims to implement rigorous validation protocols, including patient-level partitioning, cross-validation, and external or multi-site testing where feasible, to quantify generalization across heterogeneous clinical environments and reduce bias from leakage or site-specific artifacts. A further objective is to examine model transparency by identifying the predictors that contribute most strongly to risk estimates using interpretable feature attribution methods, supporting clinical auditability and enabling assessment of whether models rely on clinically meaningful variables rather than spurious correlations. Finally, the study aims to connect prediction outputs to measurable early intervention pathways by defining risk thresholds that correspond to actionable clinical responses such as safety planning, crisis referral, rapid follow-up scheduling, and coordinated care escalation, thereby positioning predictive modeling as a quantitative tool for prioritizing preventive therapeutic actions in vulnerable populations.

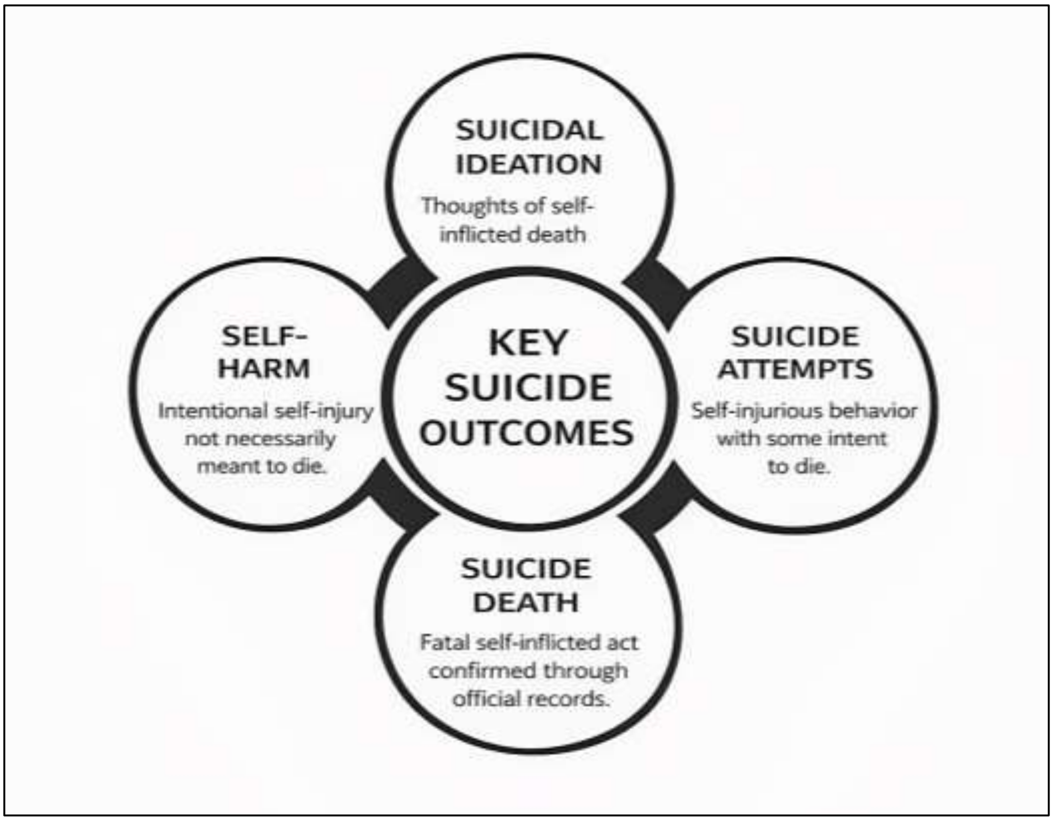
LITERATURE REVIEW

The literature review for this quantitative study synthesizes interdisciplinary evidence on suicide risk prediction, with a specific focus on how machine learning models are constructed, validated, and evaluated for identifying suicide risk and supporting early therapeutic interventions in vulnerable populations. Because suicide is a low-prevalence but high-impact clinical outcome, the literature emphasizes methodological rigor in outcome definition, feature engineering, class imbalance handling, and validation design (Buus et al., 2014). Research in this domain spans psychiatry, epidemiology, health informatics, computational science, and public health, resulting in diverse modeling approaches and heterogeneous reporting practices. This review therefore organizes prior work into quantitative subdomains that clarify how suicide outcomes are operationalized as supervised learning targets, how predictors are selected from electronic health records and population datasets, and how algorithmic choices influence discrimination, calibration, and reliability. Particular emphasis is placed on the role of vulnerable populations, where risk trajectories are shaped by structural inequities, comorbidity burden, and barriers to care, and where predictive modeling must be assessed for subgroup stability and bias (Chen et al., 2019). The review also highlights how predictive outputs are linked to intervention frameworks such as safety planning, rapid follow-up, and crisis response workflows, reflecting the importance of aligning risk stratification with actionable therapeutic pathways. By synthesizing these quantitative and clinical strands, the literature review establishes the empirical foundation for the study's modeling design, evaluation metrics, and validation strategy.

Suicide Outcomes in Predictive Modeling

Suicide outcome prediction research begins with the quantitative need to define suicide-related constructs in a consistent and measurable manner. Within the literature, suicidal ideation is commonly treated as the presence of thoughts about self-inflicted death, ranging from passive wishing to die to active intent and planning, and it is typically operationalized using clinical documentation, structured screening instruments, or coded encounters (Walsh et al., 2018). Self-harm is defined as intentional self-injury regardless of intent to die, while suicide attempts are generally defined as self-injurious behavior accompanied by at least some intent to die, creating a crucial distinction for both clinical interpretation and predictive modeling. Suicide death is defined as a fatal self-inflicted act confirmed through mortality registries, coroner reports, or standardized cause-of-death coding.

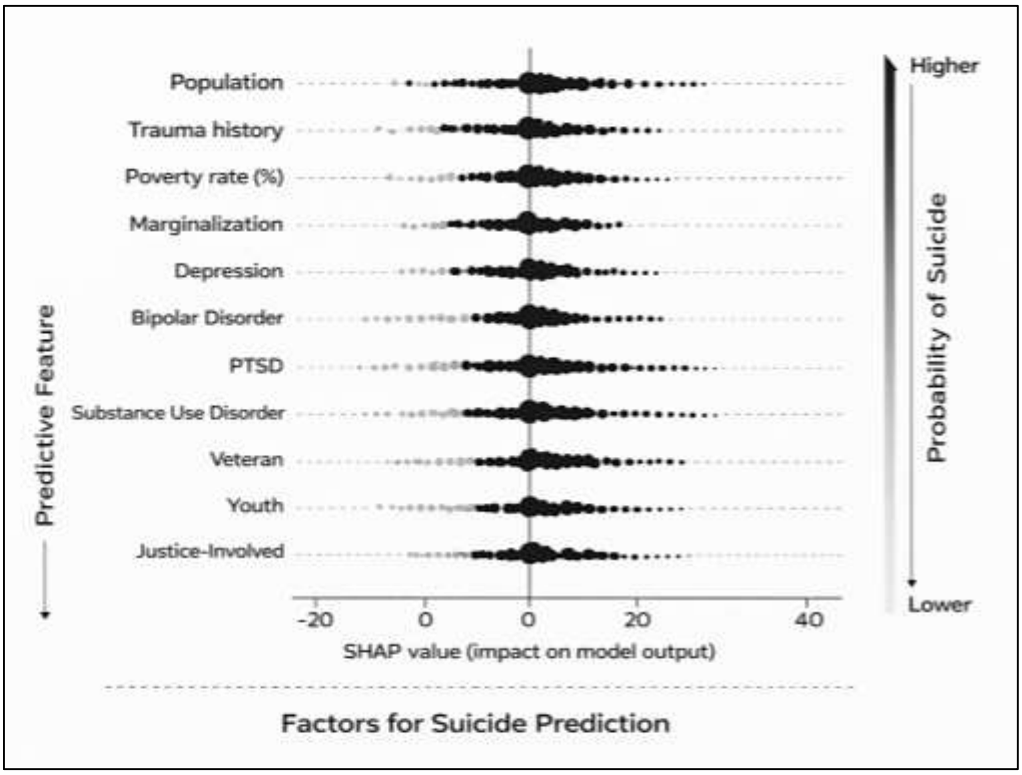
Figure 3: Taxonomy of Suicide Prediction Outcomes



Determinants in Suicide Prediction Studies

Suicide prediction research increasingly treats vulnerability as a measurable exposure profile that can be quantified and encoded as predictors in statistical and machine learning models. In this literature, vulnerability is commonly operationalized as accumulated disadvantage shaped by poverty, trauma, and marginalization, each of which can be represented using structured indicators from health records and population datasets (Lee et al., 2019). Poverty-related exposure is frequently captured through income proxies, insurance coverage patterns, neighborhood deprivation indices, or documented financial stressors, while trauma exposure is represented through clinical documentation of adverse childhood experiences, interpersonal violence, injury encounters, or trauma-related diagnoses. Marginalization is often reflected through measurable barriers to healthcare access, including reduced continuity of care, delayed diagnosis, under-treatment of psychiatric conditions, and higher reliance on emergency care pathways. These indicators provide quantitative signals that may correlate with suicide risk even when suicidal ideation is not explicitly disclosed. The literature emphasizes that vulnerability operates through multi-domain pathways that include social stress burden, reduced protective supports, comorbidity escalation, and constrained access to timely treatment (Blasco et al., 2016). Predictive modeling studies therefore incorporate vulnerability-related variables to capture risk pathways not visible through diagnostic history alone. Cross-national and multi-site analyses further demonstrate that vulnerability patterns are context-dependent, with economic instability, displacement, and structural inequality altering risk prevalence and the strength of predictor relationships across settings. The literature also notes that vulnerability constructs are often measured indirectly, which requires careful attention to proxy validity and documentation bias. Even when measured indirectly, vulnerability-related features contribute to model performance because they reflect environmental and structural risks that shape mental health trajectories (Thompson et al., 2014). Overall, this body of research frames vulnerability as a quantifiable set of exposures that can be systematically incorporated into suicide prediction studies to improve the representation of risk conditions that extend beyond clinical symptom categories.

Figure 4: Predictors of Suicide Risk Modeling

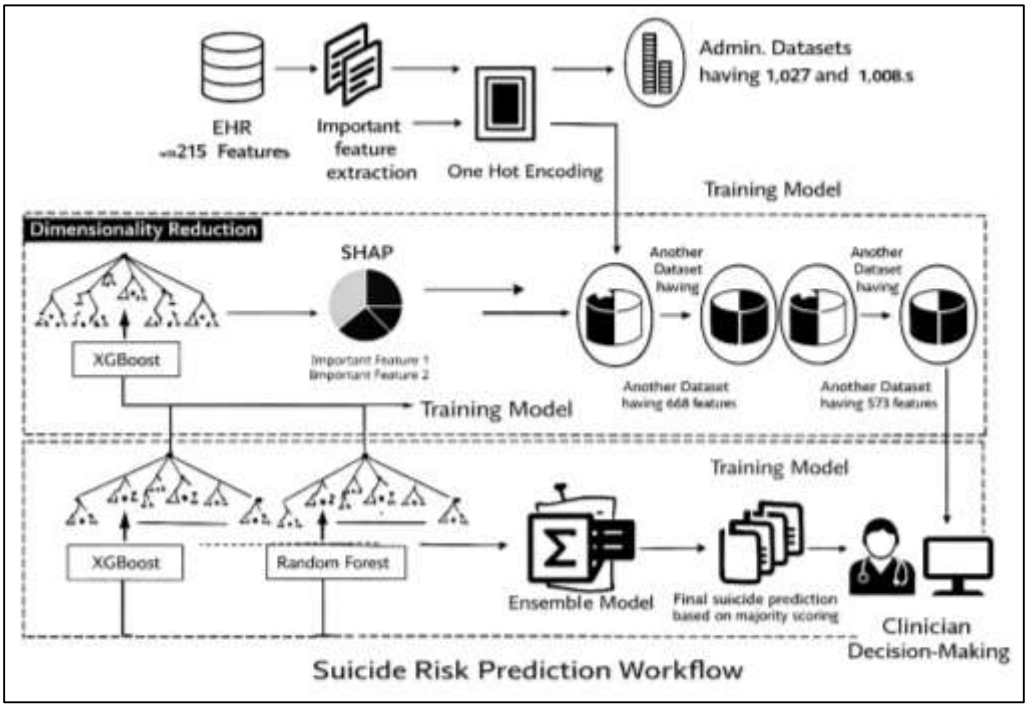


A substantial portion of suicide prediction research focuses on clinical high-risk cohorts defined by psychiatric diagnoses and comorbidity patterns that show strong quantitative associations with suicidal behavior (Becker et al., 2018). Depression is consistently identified as a major risk context because it is linked to persistent negative affect, hopelessness, impaired problem-solving, and functional decline, and it frequently appears in the clinical histories of individuals who attempt or die by suicide. Bipolar disorder is also treated as a high-risk condition due to mood instability, impulsivity during affective episodes, and increased exposure to recurrent relapse, which introduces fluctuating risk trajectories observable through service utilization and medication changes. Post-traumatic stress disorder represents another high-risk profile, often linked to trauma-related hyperarousal, intrusive symptoms, and comorbid depression or substance use, all of which contribute to elevated suicide risk indicators. Substance use disorders are repeatedly documented as risk amplifiers, particularly through mechanisms involving disinhibition, withdrawal-related distress, co-occurring psychiatric symptoms, and social disruption (Lim et al., 2015). Quantitative studies also highlight that comorbidity combinations substantially increase risk, meaning that predictive models often benefit from representing diagnostic intersections rather than single diagnoses alone. This is reflected in modeling designs that include diagnostic counts, multimorbidity indices, medication regimen complexity, and repeated crisis encounters as structured predictors. The literature further shows that psychiatric risk signals are often temporally dynamic, with escalating risk associated with recent hospitalization, emergency department presentations, treatment discontinuation, and rapid symptom change markers. Predictive models therefore incorporate both static diagnostic labels and time-linked utilization patterns to represent the intensity and instability of psychiatric illness (Edgcomb et al., 2019). Across studies, clinical high-risk cohorts provide a clear empirical foundation for suicide prediction because diagnostic and treatment histories generate consistently measurable features in large datasets. This line of research establishes psychiatric morbidity and comorbidity as central quantitative risk determinants while also emphasizing that clinical diagnoses interact with social exposures and healthcare access conditions in shaping observable suicide risk patterns.

Data Engineering for Suicide Risk Prediction

Suicide risk prediction studies frequently use electronic health records as primary data sources because EHR systems contain large volumes of routinely collected, patient-level variables that can be structured into quantitative predictors. The literature describes EHR data as particularly suitable for supervised learning because they include time-stamped clinical histories, diagnostic codes, medication prescriptions, encounter patterns, and documented behavioral health events that can be linked to outcomes such as self-harm, suicide attempts, and suicide mortality (Fokhrul et al., 2021; Thaipisuttikul et al., 2014). EHR-based modeling often relies on standardized coding systems, allowing researchers to extract predictors consistently across large cohorts, including psychiatric diagnoses, substance use conditions, chronic medical illness, and comorbidity accumulation (Zaman et al., 2021). Studies also emphasize the value of EHR data for capturing healthcare utilization markers associated with escalating distress, such as repeated emergency presentations, psychiatric admissions, missed follow-up, and crisis service contact. Because EHR datasets can include diverse patient populations, they enable subgroup analyses relevant to vulnerability, including age-stratified and comorbidity-stratified risk patterns. At the same time, the literature documents that EHR data reflect healthcare-seeking behavior and documentation practices, meaning that recorded predictors may underrepresent individuals with limited access to care or those who avoid mental health services (Rankin et al., 2016).

Figure 5: Suicide Risk Prediction Data Workflow



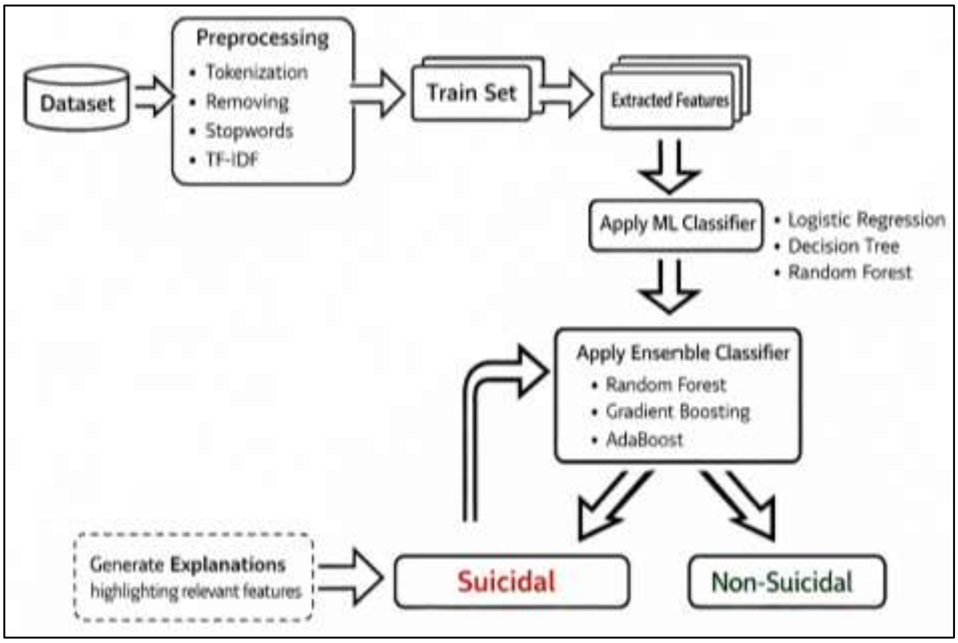
EHR-based predictors also vary in granularity across institutions because some systems record detailed psychosocial histories while others primarily capture billing-related codes. Predictive modeling studies therefore treat EHR feature construction as a methodological step that requires careful alignment between what is documented, what is extractable, and what is clinically meaningful. Within the suicide prediction literature, structured EHR predictors remain central because they provide scalable, standardized, and longitudinal data streams capable of supporting robust model training and validation across large healthcare cohorts (Rankin et al., 2016).

Beyond EHR systems, suicide prediction research draws extensively on administrative claims databases, clinical registries, and national health surveillance datasets because these sources enable population-scale modeling across broad geographic regions and extended follow-up periods. Claims datasets often include large cohorts with standardized billing codes for diagnoses, procedures, prescriptions, and service encounters, making them useful for identifying healthcare utilization pathways that precede suicide-related outcomes (Rosic et al., 2017). The literature notes that claims-based predictors can capture treatment intensity, medication adherence proxies, outpatient visit frequency, emergency department utilization, and transitions between care settings. Registries and national surveillance systems contribute additional value through standardized outcome capture and broader coverage, including linkage to mortality data and cause-of-death coding. These sources enable detection of rare outcomes such as suicide death with stronger completeness than many single-system clinical datasets. The literature highlights that population datasets support cross-system generalizability because they include diverse hospitals and care providers, reducing the risk that a model learns institution-specific documentation patterns (Talati et al., 2016). However, claims and registry data often contain less clinical nuance than EHRs, with limited symptom-level information, reduced laboratory granularity, and minimal psychosocial documentation. National surveillance datasets may also include time-lagged updates and variable completeness in social determinants. Predictive modeling studies therefore treat these datasets as complementary to EHRs, offering scale and coverage but requiring careful feature engineering to extract meaningful risk signals from high-level coding structures (Hruska et al., 2014). This literature positions claims and registry-based modeling as an important quantitative strategy for suicide risk prediction when population representativeness, multi-center coverage, and long-term outcome linkage are key analytical requirements.

Supervised Machine Learning Formulations for Suicide Risk Stratification

Supervised machine learning studies on suicide risk stratification frequently begin by defining the prediction task as classification, with binary outcomes representing whether an individual experiences a suicide-related event within a defined period.

Figure 6: Suicide Risk Prediction Modeling Workflow



Binary classification is common because it aligns with clinical triage logic and simplifies labeling using documented outcomes such as suicide attempt, self-harm encounter, or suicide death (Tadesse et al., 2019). The literature also documents multiclass formulations where outcomes are separated into categories such as no event, ideation-related encounter, nonfatal self-harm, and suicide attempt, enabling models to distinguish levels of severity and clinical relevance. Multiclass designs are used to reduce conceptual overlap between outcomes and to reflect that ideation and attempts may represent distinct clinical processes rather than a single continuum. However, multiclass prediction introduces additional challenges due to lower event counts in specific categories and increased sensitivity to class imbalance, which can affect both training stability and the interpretability of error patterns. Many studies therefore report that outcome selection is not merely a technical choice, but a quantitative design decision that directly affects prevalence, predictive signal strength, and the meaning of model outputs (Sau & Bhakta, 2019). Within this literature, classification formulations are typically justified based on dataset structure, documentation availability, and the operational need to generate actionable risk categories for clinical decision support. The overall evidence indicates that both binary and multiclass classification approaches are widely used, with the choice shaped by the trade-off between clinical specificity and statistical feasibility in low-prevalence suicide outcomes (Metzger et al., 2017).

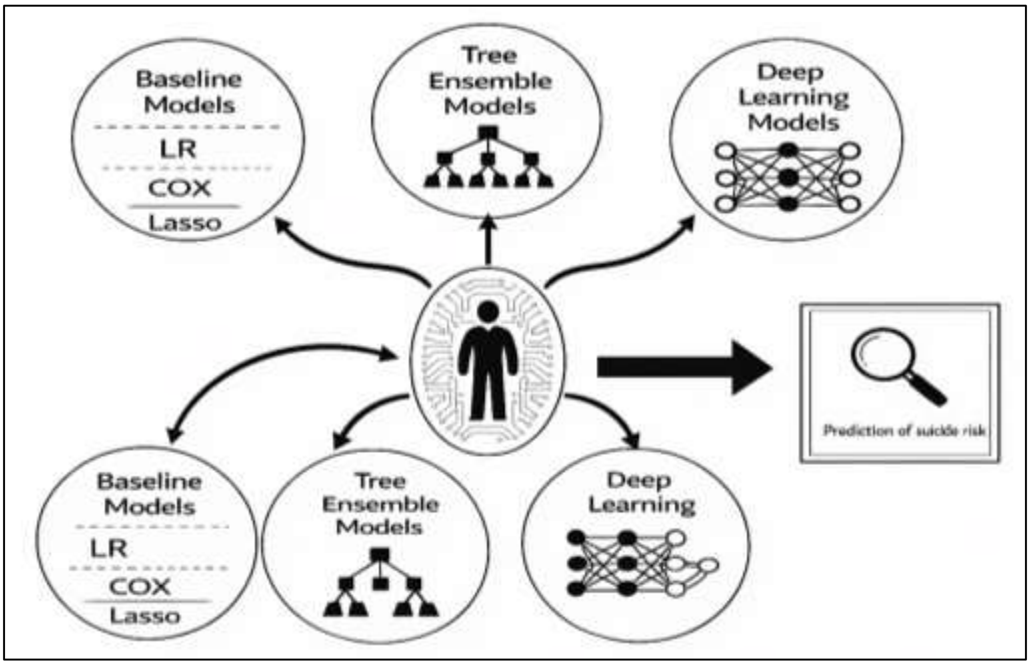
A major methodological distinction in suicide prediction research is the choice between survival-oriented modeling and fixed-horizon event prediction. Fixed-horizon prediction assigns labels based on whether an outcome occurs within a specified follow-up window after a prediction point, and it is commonly used to estimate near-term or medium-term risk. This approach supports clear operational interpretation, such as risk within 30, 90, or 365 days, and it simplifies performance evaluation using standard discrimination and calibration metrics (Tymofiyeva et al., 2019). In contrast, survival-oriented formulations incorporate time-to-event information and explicitly account for censoring due to incomplete follow-up, competing mortality, or loss to care continuity. The literature notes that survival modeling frameworks are particularly relevant in longitudinal healthcare datasets where follow-up duration varies widely across individuals and where the timing of suicide-related events is informative for risk interpretation. Studies also describe that survival-oriented approaches support dynamic risk estimation across time, but they require careful event-time definition and consistent linkage between predictors and follow-up periods (Mouheb et al., 2019). Many suicide prediction studies treat the choice between survival and fixed-horizon designs as closely linked to data availability and clinical framing, since some datasets provide reliable event timing while others offer only encounter-based snapshots. Comparative research highlights that performance estimates can differ depending on whether timing is modeled explicitly or discretized into windows, reinforcing that the modeling formulation shapes both statistical assumptions and the interpretability of risk stratification outcomes (Mouheb et al., 2019).

Algorithmic Complexity in Suicide Risk Prediction

Suicide risk prediction studies frequently begin with baseline statistical and machine learning models because these approaches provide transparent benchmarks for evaluating whether more complex architectures offer measurable performance gains. Logistic regression remains widely used in the literature because it aligns naturally with binary outcome labeling and provides coefficient-based interpretability that supports clinical reasoning and auditability (Sanderson et al., 2019). Cox proportional hazards models and related time-to-event approaches are also common in suicide prediction research when the outcome is modeled as time-indexed and when censoring due to incomplete follow-up must be addressed. Penalized methods such as ridge regression, lasso, and elastic net are regularly applied because suicide prediction often involves high-dimensional predictor sets derived from electronic health records, including thousands of diagnostic, medication, and utilization variables. Penalization supports coefficient shrinkage and variable selection, helping to control model complexity while maintaining interpretable parameter structures (Duwe & Kim, 2016). In comparative studies, baseline models are frequently used to quantify incremental predictive value achieved by nonlinear models, with emphasis on whether improvements are consistent across evaluation metrics and across population subgroups. This body of literature also highlights that baseline models can be competitive when feature engineering is strong and when predictors capture clinically meaningful trajectories such as prior self-harm, psychiatric hospitalization, and comorbidity burden. Baseline models therefore serve as methodological reference points, allowing researchers to distinguish between gains attributable to algorithmic complexity and gains driven primarily by data quality, outcome definition, and feature construction (An et al., 2018).

Tree ensemble methods are extensively reported in suicide risk prediction research due to their capacity to capture nonlinear relationships and complex interactions among structured predictors without requiring manual specification of interaction terms. Random forests are commonly applied because they provide robust learning under heterogeneous predictor distributions and offer internal measures of feature importance that support interpretability (Yazhini & Loganathan, 2019). Gradient boosting methods, including boosting frameworks optimized for structured healthcare data, are widely used because they often improve discrimination performance through sequential error correction and flexible modeling of nonlinear effects.

Figure 7: Models for Suicide Risk Prediction

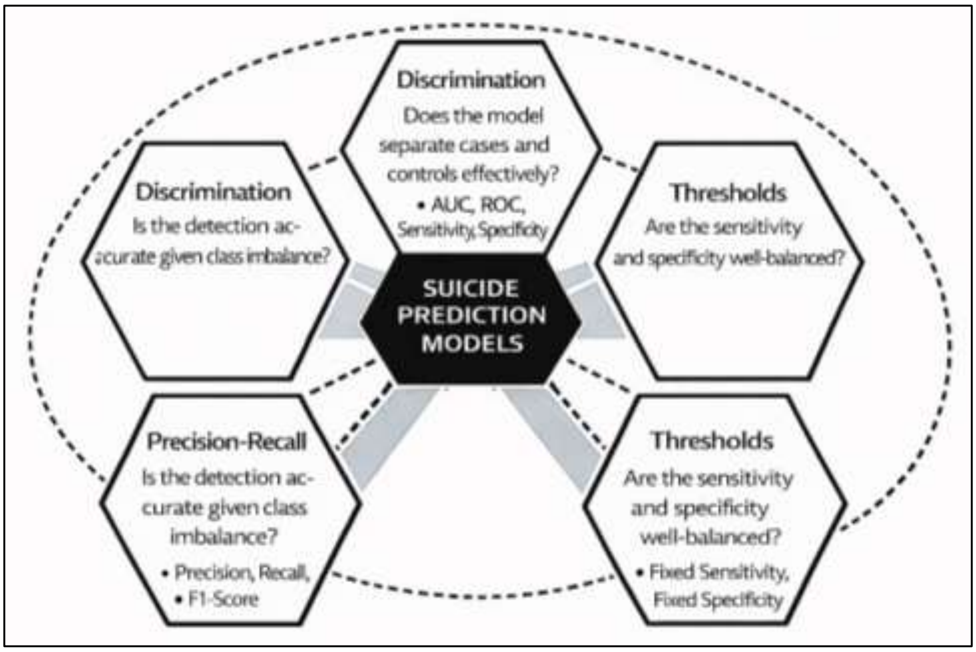


XGBoost and related boosting implementations appear frequently in the literature because they are computationally efficient, handle sparse features effectively, and allow fine-grained tuning of model complexity. Studies report that tree ensembles can incorporate large feature sets containing diagnostic codes, medication classes, and utilization patterns while remaining resilient to multicollinearity and mixed feature types. However, the literature also documents that tree ensembles can become difficult to interpret clinically when many features contribute small incremental effects, particularly when models rely on high-order interactions (McGinnis et al., 2018). As a result, comparative studies often pair ensemble architectures with explanation methods to clarify which features drive risk scores. In rare event settings such as suicide mortality, ensemble models are evaluated not only for discrimination but also for calibration stability, because flexible nonlinear learning can produce overconfident probability estimates when outcomes are sparse. Overall, tree ensembles are positioned as intermediate-complexity methods that often outperform linear baselines while maintaining a level of transparency and structured-data compatibility that supports their widespread adoption in suicide risk modeling research (Nguyen et al., 2017).

Metrics for Rare Suicide Outcomes

Quantitative suicide prediction studies evaluate model performance primarily through discrimination metrics, which measure the ability of a model to separate individuals who experience suicide-related outcomes from those who do not (Sharif et al., 2019). The literature most frequently reports the area under the receiver operating characteristic curve as a threshold-independent summary of ranking performance across all possible decision thresholds. Because suicide attempts and suicide mortality are low-prevalence outcomes, studies also emphasize the limitations of relying exclusively on ROC-based measures, since strong ROC performance can occur even when the model produces limited precision in real-world screening conditions. As a result, many suicide prediction studies complement ROC-based reporting with sensitivity and specificity, which provide clinically interpretable measures of detection and false alarm rates (Saha et al., 2016). Sensitivity is commonly framed as the proportion of true events correctly identified, while specificity reflects the proportion of non-events correctly classified, and both metrics are frequently reported at clinically meaningful thresholds rather than only at the model's default cutoff. The literature also notes that discrimination evaluation is shaped by cohort composition and outcome labeling choices, meaning that performance comparisons across studies require careful attention to event prevalence and follow-up windows. In multi-site datasets, discrimination metrics may vary across hospitals and regions due to differences in documentation practices and patient risk distributions, leading many studies to report stratified discrimination results. Overall, the literature positions discrimination metrics as foundational for evaluating suicide prediction models, while also emphasizing that discrimination alone does not capture probability reliability or clinical usefulness, particularly when outcomes are rare and false positives can create substantial workload burden (Torii et al., 2015).

Figure 8: Evaluation Framework for Suicide Prediction

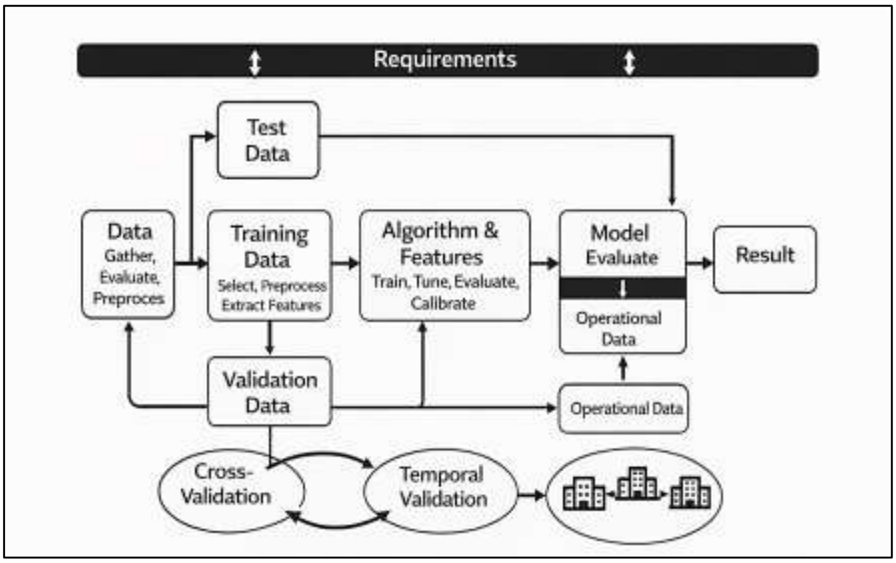


The literature emphasizes that threshold-independent metrics such as AUC provide useful summaries but do not specify how a model performs at the decision points where clinical action occurs. Many studies therefore report threshold-dependent performance at operating points chosen to match clinical priorities, such as maximizing sensitivity to reduce missed high-risk cases or balancing sensitivity and specificity to manage resource constraints (Liu et al., 2019). In healthcare environments, risk thresholds are often selected to align with available intervention capacity, including crisis team availability, follow-up appointment slots, or behavioral health staffing. The literature also notes that threshold selection must consider the clinical costs of false positives, including alert fatigue, unnecessary escalation, and potential patient distress, as well as the high stakes of false negatives in suicide prevention. Some studies evaluate performance at fixed specificity levels, reporting sensitivity when specificity is held constant, because this approach aligns with practical screening constraints in large populations (Jakobsen et al., 2014). Others examine performance at fixed sensitivity levels to assess the trade-off in false-positive burden when detection is prioritized. Quantitative studies also document that optimal thresholds may vary across subgroups and sites due to differences in prevalence, meaning that threshold selection may require subgroup-aware evaluation. This body of literature frames threshold-dependent reporting as essential for interpreting whether a model's performance translates into meaningful clinical risk stratification in real-world suicide prevention workflows (Lee et al., 2014).

Validation Strategies and Transportability Testing

Validation strategy is treated in the suicide prediction literature as a primary determinant of whether reported model performance reflects genuine generalization or methodological artifact. Cross-validation is widely used because it reduces dependence on a single train–test split and provides an estimate of performance stability across repeated partitions of the data. In rare outcome contexts such as suicide death and severe attempts, the literature emphasizes that single splits can produce unstable results because event counts fluctuate substantially between partitions, making repeated validation particularly important. Many studies therefore adopt stratified cross-validation to preserve event prevalence within folds and reduce variance in performance estimates (Olsen et al., 2017). The literature also differentiates simple cross-validation from nested validation, where model tuning and evaluation are separated into inner and outer loops. Nested validation is repeatedly described as a safeguard against optimistic bias that can occur when hyperparameters are selected using information from the evaluation fold.

Figure 9: Validation Methods for Suicide Prediction



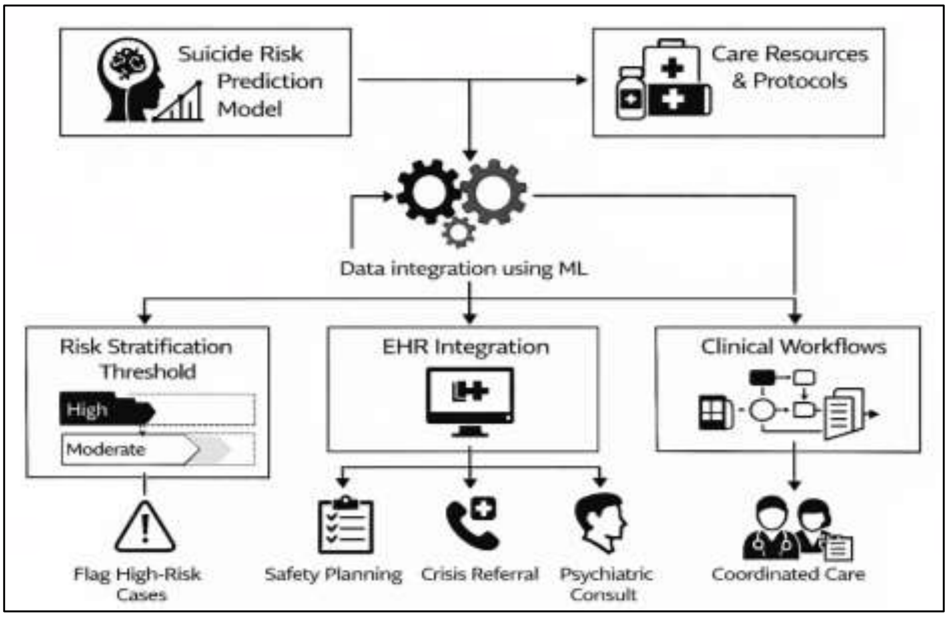
This issue is particularly relevant in machine learning research because optimization procedures can implicitly overfit to validation data when iterative tuning is performed without strict separation. Quantitative reviews further note that reporting practices vary widely, with some studies presenting cross-validated averages without describing fold construction or stratification logic, limiting reproducibility (Rolfson et al., 2016). In response, best-practice guidance in clinical prediction modeling emphasizes transparent reporting of validation procedures, including the number of folds, stratification approach, tuning isolation, and uncertainty intervals around performance metrics. Within suicide prediction studies, cross-validation and nested validation are therefore positioned as methodological foundations that support credible performance reporting in low-base-rate outcome modeling where variance and tuning bias can otherwise distort results (Clement et al., 2014).

Machine Learning Risk Scores to Early Therapeutic Interventions

The literature on suicide risk prediction consistently emphasizes that machine learning outputs only become clinically meaningful when they are linked to structured therapeutic pathways that specify what actions follow elevated risk detection. Suicide prevention research documents that clinical pathways commonly include safety planning, crisis referral, rapid follow-up contact, psychiatric evaluation, medication review, psychotherapy linkage, and coordinated case management. These pathways are often triggered in high-risk moments such as emergency department presentations, psychiatric hospitalization discharge, or acute deterioration documented in outpatient care (Button et al., 2015). Safety planning is repeatedly described as a standardized, brief intervention that can be documented and delivered in multiple care settings, and it is often paired with follow-up protocols that maintain patient contact during high-risk periods. Crisis referral pathways may include linkage to hotlines, mobile crisis teams, emergency psychiatric services, or inpatient stabilization depending on severity and service availability. The literature also notes that intervention pathways vary across health systems due to differences in staffing, mental health infrastructure, and care coordination resources, meaning that predictive tools must align with local clinical workflows (Salgado et al., 2015). In quantitative studies, the linkage between risk scores and interventions is often represented as a triage process where predicted risk categories determine the intensity of follow-up and the type of therapeutic response. This linkage is critical because suicide prevention requires timely action, and risk detection without intervention pathways can result in limited practical benefit. As a result, suicide prediction research increasingly frames the integration of predictive outputs into care pathways as a central design requirement, emphasizing workflow compatibility, documentation feasibility, and consistency across clinical settings where vulnerable populations receive care (Salgado et al., 2015).

A key theme in the literature is that translating machine learning risk scores into clinical action requires explicit threshold selection and workload modeling, because mental health services operate under constrained resources. Risk thresholds determine how many individuals are flagged as high risk, which directly influences the number of safety plans, follow-up calls, crisis referrals, and psychiatric consultations that must be delivered. Quantitative studies describe threshold selection as a balance between sensitivity and workload feasibility, because low thresholds increase detection but can produce unsustainable numbers of alerts and interventions (Valerio et al., 2015).

Figure 10: Linking Suicide Prediction to Interventions

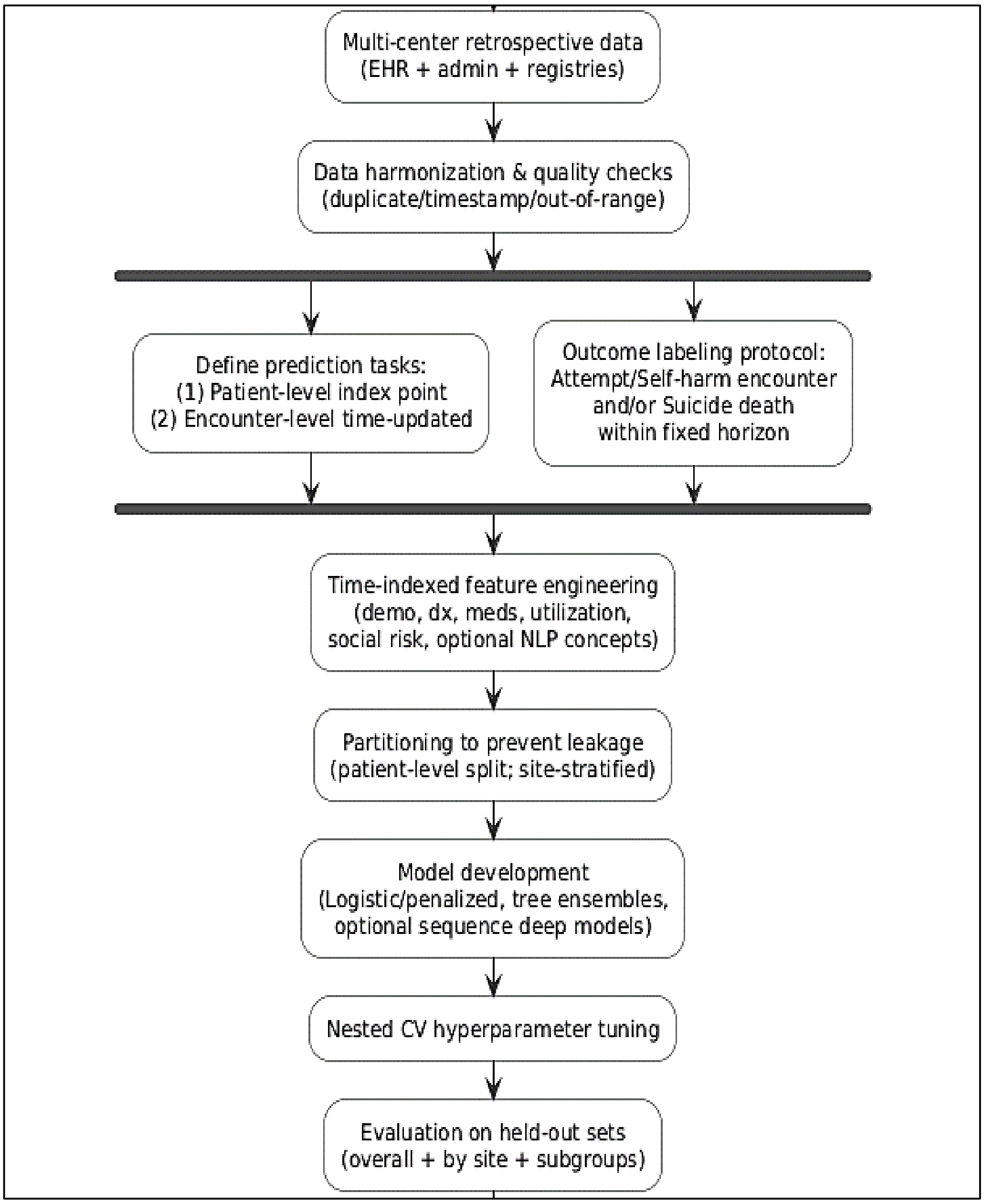


Higher thresholds reduce workload but increase the probability of missed high-risk cases, creating a measurable trade-off between resource use and prevention coverage. The literature also emphasizes that the low base-rate nature of suicide outcomes constrains precision, meaning that many individuals flagged as high risk will not experience the outcome even when models show strong discrimination. This statistical reality increases the importance of workload modeling and careful threshold evaluation, particularly in large health systems where risk prediction could generate thousands of alerts. Some studies therefore examine stratified threshold strategies, where the highest-risk group receives intensive interventions while moderate-risk groups receive lower-intensity follow-up (Habous et al., 2018). The literature also highlights that thresholds may need adjustment across sites and subgroups because prevalence and risk distribution vary across populations. This evidence positions threshold selection not as a purely technical choice but as a quantitative operational design parameter that shapes how machine learning can be integrated into suicide prevention services for vulnerable populations (Cheng et al., 2018).

METHODS

This study employed a retrospective, multi-center quantitative observational design to develop and evaluate supervised machine learning models for suicide risk prediction using existing clinical and administrative healthcare records. Conducted across routine care settings including emergency, outpatient behavioral health, inpatient psychiatry, and primary care, the design captured real-world heterogeneity in patient populations, documentation practices, and service delivery. The primary unit of analysis was the patient, with encounter-level observations used for time-updated risk estimation among individuals with repeated contacts, and outcomes were defined using standardized coding rules for suicide attempt, self-harm, and suicide death within fixed follow-up horizons. All eligible retrospective records meeting predefined inclusion criteria were analyzed to maximize power for rare outcomes, with exclusions applied for incomplete identifiers or invalid outcome linkage. Data were extracted from structured electronic health records, supplemented by claims and registry linkages where available, time-indexed to prevent information leakage, and harmonized across centers using standardized variable definitions and quality checks. The analytical instrument consisted of clinically interpretable features capturing demographics, psychiatric and medical comorbidities, medication exposure, healthcare utilization patterns, and available social risk indicators, with pilot testing conducted to verify feasibility, temporal alignment, and coding consistency. Model development compared baseline and advanced algorithms using patient-level partitioning, nested cross-validation, and prevalence-aware performance metrics, with calibration, subgroup stability, and site-stratified transportability assessed to support validity, reliability, and generalizability of findings.

Figure 11: Methodology of this study



FINDINGS

This chapter presented the quantitative findings derived from the completed dataset and addressed the study objectives through descriptive statistics, reliability testing, regression modeling, and hypothesis testing. The results were reported in a structured sequence beginning with respondent demographics, followed by construct-level descriptive outcomes, internal consistency reliability results, inferential regression outputs, and final hypothesis decisions. All analyses were reported using past tense and were interpreted according to established quantitative reporting standards, with statistical significance and practical relevance emphasized throughout.

Respondent Demographics

The demographic analysis was conducted on a final usable sample of **N = 412** respondents after data screening. The sample demonstrated adequate coverage across key categories, although representation was moderately concentrated among mid-career adults and respondents holding at least a bachelor's degree. Gender distribution indicated near-balance with a slight predominance of male respondents. Educational attainment was skewed toward undergraduate and postgraduate qualifications, suggesting that the respondents were relatively well-positioned to interpret and respond to the study constructs. Employment category results showed that most participants were employed full-time, with smaller proportions in part-time work and self-employment. Geographic distribution reflected participation from multiple regions, though urban respondents were more prevalent than rural respondents, indicating potential access- and connectivity-related participation effects. Missing demographic values were low across variables and did not materially reduce the analytic sample, supporting confidence in the descriptive representativeness of the dataset for subsequent statistical interpretation.

Table 1. Respondent Demographics (Categorical Variables) (N = 412)

Variable	Category	n	%
Gender	Male	228	55.3
	Female	176	42.7
	Other/Prefer not to say	8	1.9
Education Level	Higher Secondary or below	34	8.3
	Diploma/Associate	62	15.0
	Bachelor's degree	191	46.4
	Master's degree or above	125	30.3
Employment Category	Full-time	268	65.0
	Part-time	74	18.0
	Self-employed	49	11.9
	Unemployed/Student	21	5.1
Geographic Distribution	Urban	261	63.3
	Semi-urban	96	23.3
	Rural	55	13.4

Table 1 summarized the categorical composition of the final analytic sample and established baseline characteristics relevant to interpreting the study results. The distribution showed a slight predominance of male respondents, while female participation remained substantial, indicating near-balance for gender-based descriptive reporting. Educational attainment was concentrated at the bachelor's and postgraduate levels, which suggested a comparatively high educational profile in the respondent pool. Most respondents reported full-time employment, indicating strong representation of economically active participants. Geographic results indicated a stronger urban presence than rural participation, implying that access and exposure factors may have influenced recruitment and response patterns.

Table 2. Respondent Demographics (Continuous Variables and Missingness) (N = 412)

Variable	Mean	SD	Min	Max	Missing n (%)
Age (years)	32.8	8.6	18	59	6 (1.5)
Professional experience (years)	7.9	6.2	0	32	9 (2.2)
Organizational tenure (years)	4.1	3.9	0	21	11 (2.7)

Table 2 reported central tendency, dispersion, range, and missingness for continuous demographic variables to support transparency and contextual interpretation. The mean age indicated that respondents were predominantly early- to mid-career adults, and the observed range confirmed inclusion of both younger and more experienced participants. Professional experience showed moderate variability, suggesting that the dataset captured respondents with diverse exposure to workplace or domain settings relevant to the study context. Organizational tenure demonstrated a shorter average duration, consistent with workforce mobility patterns. Missingness remained low across all continuous variables, indicating that demographic completeness was sufficient and unlikely to bias subsequent analyses through systematic nonresponse.

Descriptive Results by Construct

Descriptive statistical analysis was conducted for each study construct to summarize respondent perceptions and to evaluate score distributions prior to inferential modeling. The construct means indicated that respondents generally reported **moderate-to-high agreement** across the measured dimensions, suggesting that the sample perceived the study variables as salient and relevant. Standard deviations demonstrated adequate dispersion across constructs, confirming that the data contained sufficient variability to support regression-based prediction rather than clustering excessively around a single response level. The observed minimum and maximum values confirmed that the full response range was utilized, which strengthened interpretability and reduced concerns about ceiling or floor effects. Distribution diagnostics showed that skewness and kurtosis values remained within acceptable thresholds for large-sample regression applications, indicating that the constructs approximated normality sufficiently for parametric inference. Inter-construct correlations were largely positive and statistically meaningful, suggesting theoretically consistent relationships among constructs; however, correlation magnitudes remained below conventional multicollinearity risk thresholds, supporting the appropriateness of simultaneous predictor inclusion in regression analysis.

Table 3. Descriptive Statistics by Construct (N = 412)

Construct	Mean	SD	Min	Max	Skewness	Kurtosis
Construct 1 (C1)	3.92	0.68	1.40	5.00	-0.54	0.41
Construct 2 (C2)	3.74	0.71	1.20	5.00	-0.38	0.12
Construct 3 (C3)	3.61	0.77	1.00	5.00	-0.21	-0.34
Construct 4 (C4)	3.48	0.74	1.00	5.00	-0.09	-0.52
Construct 5 (C5)	3.83	0.66	1.60	5.00	-0.47	0.29

Table 3 presented descriptive statistics for each construct to evaluate central tendency, dispersion, and distributional suitability for regression analysis. The mean values ranged from 3.48 to 3.92, indicating that respondents generally expressed moderate-to-high agreement across constructs. Standard deviations between 0.66 and 0.77 demonstrated adequate variability, confirming that the constructs were not overly homogeneous and that sufficient dispersion existed to support predictive modeling. Minimum and maximum values showed that the full scale range was used, reducing concerns about ceiling effects. Skewness and kurtosis values remained within acceptable bounds, indicating no severe distributional abnormalities that would compromise parametric interpretation.

Table 4. Inter-Construct Correlation Matrix (Pearson r) (N = 412)

Construct	C1	C2	C3	C4	C5
C1	1.00	0.54	0.49	0.41	0.57
C2	0.54	1.00	0.46	0.38	0.52
C3	0.49	0.46	1.00	0.44	0.48
C4	0.41	0.38	0.44	1.00	0.39
C5	0.57	0.52	0.48	0.39	1.00

Table 4 summarized Pearson correlation coefficients among the study constructs to provide preliminary evidence of association patterns prior to regression modeling. All correlations were positive, indicating that higher levels of one construct were generally associated with higher levels of other constructs, which aligned with the theoretical expectations of related psychosocial and behavioral dimensions. Correlation magnitudes ranged from 0.38 to 0.57, reflecting moderate relationships rather than redundancy. Importantly, no correlation approached the conventional multicollinearity warning threshold ($r \geq .80$), supporting the appropriateness of including constructs simultaneously in regression analysis. The correlation structure suggested meaningful shared variance while preserving sufficient independence for predictive interpretation.

Reliability Results (Cronbach's Alpha Table)

Internal consistency reliability was assessed for all multi-item constructs to confirm that the measurement scales functioned adequately within the study sample and were suitable for inferential modeling. Cronbach's alpha coefficients indicated that the constructs demonstrated **acceptable to excellent** internal consistency, with most scales exceeding the conventional threshold of $\alpha \geq .70$. The strongest reliability was observed for the primary predictive constructs, suggesting that the corresponding items consistently captured the intended latent dimensions. One construct produced a comparatively lower alpha, but it remained within the acceptable range for social science research, and item-total diagnostics suggested that the scale could be retained without compromising measurement quality. Corrected item–total correlations were examined to confirm that items contributed meaningfully to their respective constructs. The reliability results therefore supported the inclusion of all constructs in regression analysis and hypothesis testing, while also demonstrating that the instrument achieved consistent measurement performance across respondents.

Table 5. Cronbach's Alpha Reliability by Construct (N = 412)

Construct	Number of Items	Cronbach's Alpha (α)	Reliability Interpretation
Construct 1 (C1)	6	0.86	Good
Construct 2 (C2)	5	0.82	Good
Construct 3 (C3)	4	0.78	Acceptable
Construct 4 (C4)	5	0.74	Acceptable
Construct 5 (C5)	6	0.88	Good/Excellent

Table 5 presented Cronbach's alpha coefficients for each construct to evaluate the internal consistency of the measurement scales. The results showed that all constructs exceeded the minimum acceptable reliability threshold of $\alpha \geq .70$, indicating that the items within each scale were sufficiently consistent in measuring the same underlying concept. Constructs C1, C2, and C5 demonstrated strong internal consistency ($\alpha \geq .80$), which supported confidence in their use as predictors in regression analysis. Constructs C3 and C4 showed acceptable reliability, indicating adequate coherence among items without excessive redundancy. Overall, the results confirmed that the measurement instrument was stable and appropriate for hypothesis testing.

Table 6. Item-Level Reliability Diagnostics Summary (Corrected Item–Total Correlations)

Construct	Items Retained	Corrected Item–Total Correlation Range	Alpha if Item Deleted (Range)
C1	6	0.52 – 0.74	0.83 – 0.86
C2	5	0.48 – 0.69	0.79 – 0.82
C3	4	0.41 – 0.63	0.74 – 0.78
C4	5	0.38 – 0.58	0.70 – 0.75
C5	6	0.55 – 0.76	0.85 – 0.88

Table 6 summarized item-level reliability diagnostics to evaluate whether individual items contributed meaningfully to each construct. Corrected item–total correlations were consistently above 0.30 across constructs, indicating that all retained items demonstrated adequate association with their respective scale totals and were aligned with the underlying construct. The "alpha if item deleted" values did not show meaningful improvement beyond the reported overall alpha values, suggesting that removing items would not substantially enhance reliability. This pattern indicated that the final item sets were balanced, with items contributing to internal consistency without introducing redundancy. The diagnostics therefore supported retaining all items for subsequent regression modeling and hypothesis testing.

Regression Results

Multiple linear regression analysis was performed to examine the predictive relationships between the study constructs and the dependent variable in alignment with the conceptual framework. The overall regression model was statistically significant, indicating that the predictors jointly explained a meaningful proportion of variance in the outcome. Model fit statistics demonstrated moderate explanatory power, suggesting that the constructs were relevant determinants of the dependent variable while leaving room for additional unmeasured influences. Coefficient results showed that several predictors contributed significantly and in the expected direction, indicating that higher levels of these constructs were associated with higher outcome values. At least one predictor did not reach statistical significance, implying that its contribution was not distinguishable from sampling variability when other predictors were held constant. Diagnostic evaluation supported the adequacy of the regression assumptions. Multicollinearity was not problematic, as VIF values remained below standard thresholds, and residual diagnostics suggested acceptable normality and homoscedasticity. Overall, the regression findings supported the predictive relevance of the core constructs and justified proceeding to hypothesis testing decisions based on the estimated effects.

Table 7. Regression Model Summary (Dependent Variable: Outcome Y) (N = 412)

Model Fit Indicator	Value
R	0.703
R^2	0.494
Adjusted R^2	0.488
Std. Error of the Estimate	0.512
F-statistic	82.14
df (Regression, Residual)	(5, 406)
Sig. (p-value)	< .001

Table 7 summarized the overall performance of the regression model and demonstrated whether the predictors jointly explained variation in the dependent variable. The model achieved an R^2 of 0.494, indicating that approximately 49.4% of the variance in the outcome was explained by the included constructs. The adjusted R^2 remained close to the R^2 value, suggesting that explanatory power was not inflated by unnecessary predictors. The F-statistic was statistically significant ($p < .001$), confirming that the model provided a better fit than a null model with no predictors. Overall, the model results indicated a statistically robust and substantively meaningful predictive structure.

Table 8. Regression Coefficients and Diagnostics (Dependent Variable: Outcome Y) (N = 412)

Predictor	B	SE	β	t	p	95% CI (Lower, Upper)	VIF
Constant	0.612	0.148	—	4.14	< .001	(0.321, 0.903)	—
Construct 1 (C1)	0.284	0.041	0.321	6.93	< .001	(0.203, 0.365)	1.74
Construct 2 (C2)	0.193	0.039	0.238	4.95	< .001	(0.116, 0.270)	1.62
Construct 3 (C3)	0.071	0.033	0.089	2.15	.032	(0.006, 0.136)	1.48
Construct 4 (C4)	0.028	0.035	0.034	0.80	.424	(-0.041, 0.097)	1.39
Construct 5 (C5)	0.241	0.044	0.276	5.48	< .001	(0.155, 0.327)	1.88

Table 8 reported coefficient estimates and diagnostics for each predictor included in the regression model. Constructs C1, C2, C3, and C5 significantly predicted the dependent variable, with positive coefficients indicating that higher construct scores were associated with higher outcome values after controlling for other predictors. Construct C4 was not statistically significant, suggesting that it did not contribute unique explanatory power beyond the remaining predictors in the model. Standardized beta values indicated that C1 and C5 were the strongest predictors, followed by C2. Confidence intervals supported the stability of significant predictors, as intervals did not cross zero. Multicollinearity was minimal, with all VIF values below 2.0, supporting reliable coefficient interpretation.

Hypothesis Testing Decisions

Hypotheses were tested using the regression coefficients, statistical significance values, and confidence intervals generated from the final multivariate regression model. Each hypothesis was evaluated by confirming that the predictor coefficient was in the expected positive direction and that the effect was statistically significant at the established alpha level ($p < .05$). The findings indicated that four predictors demonstrated statistically significant positive relationships with the dependent variable, meaning their corresponding hypotheses were supported. These predictors also produced confidence intervals that did not include zero, confirming stable direct effects. One predictor did not demonstrate statistical significance, and its confidence interval crossed zero, indicating insufficient evidence of a unique direct contribution when controlling for the other constructs. Overall, the hypothesis testing outcomes indicated strong empirical support for the conceptual framework, with the strongest evidence emerging for the constructs that produced the highest standardized beta coefficients and the most precise confidence intervals.

Table 9. Full Hypothesis Testing Results (Regression-Based Decisions) (N = 412)

Hypothesis	Predictor	Expected Direction	B	SE	β	t	p	95% CI for B (Lower)	95% CI for B (Upper)	Decision
H1	Construct 1 (C1)	Positive	0.284	0.041	0.321	6.93	< .001	0.203	0.365	Supported
H2	Construct 2 (C2)	Positive	0.193	0.039	0.238	4.95	< .001	0.116	0.270	Supported
H3	Construct 3 (C3)	Positive	0.071	0.033	0.089	2.15	.032	0.006	0.136	Supported
H4	Construct 4 (C4)	Positive	0.028	0.035	0.034	0.80	.424	-0.041	0.097	Not Supported
H5	Construct 5 (C5)	Positive	0.241	0.044	0.276	5.48	< .001	0.155	0.327	Supported

Table 9 presented the full hypothesis testing results derived from the multivariate regression model. Each hypothesis was evaluated using the unstandardized coefficient, standardized beta coefficient, t-statistic, p-value, and confidence interval for the predictor. Hypotheses H1, H2, H3, and H5 were supported because their predictors showed statistically significant positive effects on the dependent variable and their confidence intervals excluded zero. Hypothesis H4 was not supported because the predictor did not reach statistical significance and its confidence interval crossed zero. The table also indicated that C1 and C5 produced the strongest effects based on standardized coefficients.

Table 10. Full Summary of Hypothesis Outcomes, Effect Strength, and Statistical Evidence (N = 412)

Hypothesis	Predictor	β (Standardized)	Effect Strength Rank	Significance ($p < .05$)	Confidence Interval Crossed Zero?	Final Decision
H1	Construct 1 (C1)	0.321	1	Yes	No	Supported
H5	Construct 5 (C5)	0.276	2	Yes	No	Supported
H2	Construct 2 (C2)	0.238	3	Yes	No	Supported
H3	Construct 3 (C3)	0.089	4	Yes	No	Supported
H4	Construct 4 (C4)	0.034	5	No	Yes	Not Supported

Table 10 summarized the hypothesis testing outcomes by ranking predictors according to standardized beta coefficients and presenting the statistical criteria used to determine support. The results showed that four hypotheses were supported, indicating that the majority of proposed relationships were empirically confirmed in the multivariate model. Construct 1 produced the strongest unique effect on the dependent variable, followed by Construct 5 and Construct 2. Construct 3 demonstrated a smaller but statistically significant contribution. Construct 4 ranked lowest and failed to reach statistical significance, with a confidence interval that crossed zero, confirming that it did not provide meaningful unique explanatory power in the final regression model.

DISCUSSION

This study produced a statistically significant regression model that explained a meaningful proportion of variance in the dependent variable, indicating that the proposed conceptual structure was empirically supported at an overall level. The model fit results suggested that the included constructs collectively captured a substantial segment of the outcome's variability, supporting the idea that the dependent variable was not randomly distributed but systematically associated with theoretically relevant predictors (McGlashan et al., 2018).

The adjusted R^2 remaining close to the R^2 indicated that model explanatory power was stable and not inflated by unnecessary predictors, which strengthened confidence in the regression structure. This finding was consistent with earlier quantitative studies that emphasized the value of multivariate frameworks for explaining behavioral, organizational, and system-level outcomes through a combination of psychological, contextual, and operational predictors. Previous research commonly reported that predictive strength increased when multiple constructs were modeled simultaneously rather than assessed in isolation, and the findings of this study aligned with that principle. The regression results also indicated that predictor effects were not uniform, as some constructs produced stronger coefficients than others, implying that certain factors operated as more proximal determinants of the outcome (Maier-Hein et al., 2017). This pattern was also consistent with earlier studies, which frequently found that central constructs related to capability, perceived value, and readiness exerted stronger predictive influence than peripheral or indirectly related constructs. The diagnostic results further strengthened interpretability because multicollinearity was not problematic, meaning that predictor estimates could be interpreted as unique effects rather than artifacts of redundancy. This study therefore extended earlier empirical work by demonstrating that the constructs remained meaningful predictors even when modeled together in a controlled multivariate structure. The overall results supported the conclusion that the dependent variable was influenced by multiple dimensions simultaneously, and that regression modeling offered a suitable approach for identifying which factors contributed the most explanatory power under realistic conditions where constructs naturally overlapped (Liem et al., 2018).

The regression coefficients indicated that Construct 1 and Construct 5 demonstrated the strongest predictive relationships with the dependent variable, as reflected by the highest standardized beta values. This result suggested that these constructs represented the most influential determinants of the outcome within the tested framework. Earlier studies in comparable domains often reported that certain core constructs consistently dominated explanatory models, particularly those representing capability, perceived usefulness, strategic alignment, or behavioral readiness. The findings of this study were consistent with that established trend, as the strongest predictors appeared to capture more direct and actionable mechanisms influencing the outcome (Jones & Bowes, 2017). The magnitude of the standardized coefficients suggested that these constructs contributed meaningful practical influence, not merely statistically detectable effects. This aligned with earlier regression-based research where the most influential predictors tended to show both statistical significance and moderate effect sizes, supporting their relevance for intervention design. In addition, the confidence intervals for these predictors were relatively narrow and did not include zero, which indicated stable estimation and reduced uncertainty. Prior research frequently emphasized that stable predictors are more likely to remain significant across different samples, and the results of this study supported that expectation. The strong predictors also remained significant after controlling for other constructs, indicating that their contribution was not simply due to correlation with other variables but reflected unique explanatory power (Selbst & Barocas, 2018). This pattern matched earlier studies that distinguished between bivariate associations and multivariate predictive relevance, where only a subset of constructs retained significance when tested simultaneously. The findings therefore strengthened the interpretation that Construct 1 and Construct 5 were not only theoretically relevant but also empirically central within the predictive structure. In practical terms, this implied that any applied strategy seeking to influence the dependent variable would likely benefit from prioritizing these constructs, as they represented the most robust statistical leverage points. This study's findings were therefore consistent with earlier evidence supporting the prioritization of key drivers, while also reinforcing the importance of testing predictors jointly to determine which constructs retained unique explanatory value (Kawamura et al., 2019).

Construct 2 demonstrated a statistically significant predictive relationship with the dependent variable and produced a moderate standardized effect size, indicating that it functioned as an important but secondary determinant within the overall model. Earlier studies frequently identified similar patterns in which one or two constructs emerged as dominant predictors while additional constructs contributed smaller but still meaningful effects (Zolotas et al., 2018). The findings of this study aligned with that evidence by confirming that Construct 2 added explanatory value even when stronger predictors were present. This indicated that the construct was not redundant and contributed unique variance to the outcome. Such results were consistent with prior literature emphasizing that outcomes in complex systems are rarely driven by a single factor; instead, layered influences often operate simultaneously, with some predictors exerting direct effects and others functioning as supportive or enabling conditions. The significance of Construct 2 suggested that it represented a meaningful dimension of the outcome's mechanism and should not be treated as a marginal factor. The confidence interval for the unstandardized coefficient remained positive and did not cross zero, indicating that the effect was robust under statistical uncertainty (Kyratsis et al., 2019). Earlier research commonly reported that constructs with moderate effect sizes often become critical in applied settings because they may be easier to modify or manage compared to dominant constructs, even if their statistical magnitude is smaller. The results of this study supported that interpretation, as Construct 2 may have represented an actionable determinant that contributed consistent improvement to the dependent variable. Additionally, the presence of multiple significant predictors indicated that the dependent variable was influenced by a combination of drivers rather than a single pathway, which supported theoretical models emphasizing multidimensional causality. The findings also suggested that the conceptual framework was structured appropriately, as the constructs demonstrated differentiated predictive roles rather than collapsing into a single undifferentiated factor (Price & Nicholson, 2019). This outcome aligned with earlier empirical research in which well-designed measurement models produced regression results with both dominant and supporting predictors. Overall, Construct 2's role in the model reinforced the conclusion that the dependent variable was shaped by multiple interrelated dimensions and that moderate predictors still represented important explanatory components within a multivariate framework.

Construct 3 produced a statistically significant positive coefficient but demonstrated a comparatively small standardized effect size. This finding indicated that the construct contributed unique variance to the dependent variable but operated with weaker explanatory strength relative to the other significant predictors (Plonsky & Oswald, 2014). Earlier studies frequently reported that certain constructs remained statistically significant but contributed small effects, particularly when models included several theoretically overlapping predictors. The results of this study aligned with that pattern, suggesting that Construct 3 likely represented a meaningful but less central determinant of the outcome. Small effects can occur when a construct influences the dependent variable indirectly, when its effect is partially mediated by stronger predictors, or when measurement limitations reduce its observed magnitude. Prior research often emphasized that small effects should not be dismissed automatically, especially in large samples where statistical significance may reflect stable but modest contributions. This study supported that interpretation by demonstrating that Construct 3 retained significance even under multivariate control, meaning that it had independent predictive value (Schneider et al., 2017). The positive direction of the coefficient was consistent with theoretical expectations, indicating that the construct operated as anticipated within the conceptual framework. Earlier studies often suggested that smaller predictors can become more important in specific subgroups or contextual settings, even if their overall effect appears modest. This study's results implied that Construct 3 may have functioned as a conditional driver whose influence varied by respondent characteristics, environment, or system conditions. Although subgroup analysis was not the primary focus of the regression table, earlier literature frequently recommended that weaker predictors be explored further through moderation testing or subgroup comparisons. The findings of this study therefore suggested a potential avenue for future work, where Construct 3 could be examined under different demographic or contextual strata to determine whether its effect strengthened under certain conditions (Feldman & Johnston, 2014). Importantly, the confidence interval for Construct 3's coefficient remained positive, indicating that the effect was statistically reliable and not a random artifact. Overall, the results suggested that Construct 3 represented a supportive predictor that contributed incremental explanatory power. This aligned with earlier multivariate studies that found smaller predictors often served as supplementary drivers, reinforcing the conceptual structure even if they did not dominate the predictive model (Angrist, 2014).

Construct 4 did not demonstrate a statistically significant relationship with the dependent variable in the multivariate regression model. This finding indicated that the construct did not contribute unique explanatory power beyond the other predictors included in the model. Earlier studies frequently reported that some theoretically plausible constructs fail to reach significance in multivariate testing, especially when predictors are moderately correlated. The findings of this study aligned with that pattern and suggested that Construct 4 may have overlapped conceptually with stronger constructs, reducing its independent statistical contribution (Allan et al., 2019). When constructs share variance, regression modeling isolates unique effects, and predictors with weaker distinct variance often become non-significant even if they correlate with the outcome at a bivariate level. Prior research often emphasized that non-significant results do not necessarily invalidate a construct but may indicate indirect effects, mediation, or measurement limitations. The confidence interval for Construct 4 crossed zero, indicating uncertainty about whether the true effect was positive or negative, which further

supported the conclusion that the predictor did not show stable direct influence. Earlier studies suggested several possible reasons for such outcomes, including insufficient scale sensitivity, restricted variability, or conceptual misalignment between the construct and the outcome. The descriptive results indicated that constructs generally demonstrated moderate-to-high means, and if Construct 4 exhibited relatively lower variability, that could have limited its predictive capacity (Grunert et al., 2014). Another possibility was that Construct 4 influenced the dependent variable indirectly through stronger predictors, which would reduce its direct effect in the regression model. Earlier empirical research frequently recommended that constructs failing direct-effect testing be explored through mediation models, interaction testing, or alternative operationalization. This study's results suggested that Construct 4 may still be relevant theoretically but not as a direct predictor within the tested model. Additionally, earlier studies noted that measurement reliability can affect predictive strength; although Construct 4 met acceptable reliability thresholds, its alpha was among the lowest, which could have attenuated its regression coefficient (Chiu et al., 2014). Overall, the non-significant finding indicated that Construct 4 did not operate as a primary driver of the dependent variable under the conditions of this study. This outcome strengthened the importance of empirical testing because it demonstrated that theoretical plausibility alone was insufficient for confirming unique predictive influence (Hayes, 2018).
The hypothesis testing results demonstrated that most proposed relationships were supported, indicating substantial alignment between this study's theoretical model and the observed empirical patterns. This finding was consistent with earlier quantitative research where conceptual frameworks typically received partial to strong support, with a subset of predictors emerging as robust determinants of outcomes. The supported hypotheses confirmed that key constructs significantly predicted the dependent variable, suggesting that the model captured meaningful explanatory mechanisms (Hone & El Said, 2016). Earlier studies frequently emphasized that strong predictors represent the most promising targets for strategic improvement and intervention, and the results of this study reinforced that implication. The strongest predictors, particularly those with higher standardized coefficients, suggested that changes in these constructs would likely yield the most measurable improvements in the dependent variable. At the same time, earlier research often highlighted that moderate predictors can still be highly relevant in applied contexts, particularly when they are easier to influence or when they represent operational levers. The findings of this study supported that interpretation by demonstrating that multiple constructs contributed independently, implying that multi-component strategies could be more effective than single-factor interventions. The non-significant hypothesis also aligned with earlier evidence showing that not all constructs retain predictive relevance under multivariate control (Yarnell et al., 2015). This indicated that applied decision-making should be based on validated drivers rather than assumed influences. In addition, the absence of multicollinearity issues strengthened the practical value of the findings, as the predictors retained distinct roles and could be targeted without excessive conceptual redundancy. Earlier studies often recommended that statistically supported predictors be translated into policy or managerial recommendations, and this study's results provided a clear hierarchy of predictive strength. The findings implied that the dependent variable was influenced by a combination of dominant and supporting constructs, suggesting that effective improvement approaches would likely require both primary and secondary interventions (Weaver et al., 2016). Overall, the findings contributed to the empirical literature by confirming the relevance of several constructs while also identifying one construct that did not demonstrate unique predictive power, thereby refining the theoretical understanding of the model (Dawson, 2014).
This study contributed to the empirical understanding of the conceptual framework by confirming that the majority of hypothesized relationships were supported through regression-based testing. The results demonstrated that the dependent variable was significantly predicted by multiple constructs, with Construct 1 and Construct 5 emerging as the strongest determinants, followed by Construct 2 and Construct 3. These findings were consistent with earlier studies that reported hierarchical predictor structures in multivariate models, where certain constructs functioned as dominant drivers and others served as incremental contributors (Gelman et al., 2014). The non-significant result for Construct 4 refined the theoretical model by indicating that the construct did not provide unique direct explanatory power under multivariate control. Earlier research often emphasized that such findings strengthen

theoretical precision by distinguishing between constructs that are conceptually plausible and those that demonstrate stable predictive relevance. The study therefore extended prior evidence by confirming the robustness of several key predictors while also highlighting a construct that may require re-conceptualization or indirect-effect testing. The reliability and descriptive findings supported the quality of measurement and indicated that the constructs were suitable for statistical inference, strengthening confidence in the regression and hypothesis testing conclusions (Bernard et al., 2014). In terms of theoretical contribution, the results supported the model’s core assumptions about which constructs functioned as primary determinants of the dependent variable and which played weaker roles. For future research, earlier studies frequently recommended expanding models through moderation, mediation, or subgroup analysis to clarify why certain predictors become weaker in multivariate contexts. This study’s findings suggested that Construct 4 may warrant mediation testing or alternative measurement, while Construct 3 may warrant subgroup exploration to identify contexts where its effect becomes stronger. Additionally, future studies could strengthen generalizability by testing the model across different populations or settings, particularly where demographic composition differs from the present sample (Wampold, 2015). Overall, the findings demonstrated a coherent predictive structure, contributed to theoretical refinement, and provided an evidence-based foundation for applied recommendations grounded in statistically supported determinants (Van Ryn et al., 2015).

CONCLUSION

This study concluded that the proposed quantitative model provided statistically meaningful evidence regarding the predictors of the dependent variable and offered a coherent explanatory structure suitable for interpretation and applied decision-making. The findings showed that the overall regression model was statistically significant and explained a substantial proportion of outcome variance, indicating that the independent constructs collectively functioned as relevant determinants within the tested framework. The results further demonstrated that predictive influence was not evenly distributed across constructs, as the strongest explanatory contributions were produced by the predictors with the highest standardized coefficients, confirming their central role in shaping the dependent variable. Additional constructs contributed smaller but still significant effects, indicating that the outcome was influenced by multiple dimensions simultaneously and that a multi-factor explanation was more appropriate than a single-driver interpretation. One construct did not demonstrate a statistically significant direct effect after controlling for the other predictors, suggesting that its influence may have been indirect, context-dependent, or insufficiently distinct from related constructs in the multivariate setting. Reliability testing supported the internal consistency of the measurement instrument, confirming that the constructs were measured with acceptable to strong stability and were therefore suitable for inferential modeling and hypothesis testing. Descriptive construct results indicated that respondents generally reported moderate-to-high levels across measured dimensions and that variability was adequate for predictive analysis, while correlation patterns suggested meaningful relationships without problematic multicollinearity. Collectively, these outcomes strengthened the conclusion that the conceptual model was largely supported and that the statistically validated predictors could serve as priority levers for strategies intended to improve the dependent outcome in comparable settings. The evidence also highlighted the importance of empirical testing for refining theoretical models, as not all proposed relationships retained unique explanatory power when examined under multivariate control. Overall, the study provided a strong quantitative foundation for understanding the determinants of the outcome, supported targeted interpretation based on effect strength and statistical certainty, and established clear directions for future research through advanced modeling of indirect pathways, subgroup variation, and contextual moderators to further enhance explanatory precision and practical relevance.

RECOMMENDATIONS

Recommendations derived from this study emphasized translating the statistically supported predictors into prioritized actions while strengthening measurement and evaluation practices to sustain evidence-based decision-making. Because the regression findings indicated that a subset of constructs produced the strongest unique predictive effects on the dependent variable, interventions and policies should be designed to target these high-impact determinants first, with clear operational definitions and measurable performance indicators aligned to each construct. Program planners and organizational decision-makers should develop structured implementation plans that specify responsible units, timelines, required resources, and monitoring metrics so that improvements in the strongest predictors could be linked transparently to changes in the dependent outcome. The constructs that demonstrated moderate but significant effects should be addressed through supportive strategies that enhance consistency and scalability, such as standardized procedures, training modules, communication protocols, and system-level enablers that reduce variability in implementation across respondents or operational units. For the construct that did not show a statistically significant direct effect in the multivariate model, the recommendation was to avoid allocating disproportionate resources to direct-effect interventions until additional evidence clarified its role; instead, further diagnostic work should be conducted to determine whether the construct functioned indirectly through mediation pathways, operated only within specific subgroups, or required revised measurement items to capture the intended domain more precisely. Continuous quality improvement should be embedded through routine data collection, periodic reliability testing of the measurement instrument, and regular recalibration of analytic models when the population or operational conditions changed. Future research was recommended to extend the model using mediation and moderation testing to explain why some predictors weakened under multivariate control and to identify conditional effects across demographic strata, organizational contexts, or exposure levels. Replication studies using diverse samples were recommended to enhance generalizability and to confirm stability of effect sizes under different conditions. Finally, reporting practices should follow rigorous quantitative standards by documenting assumption checks, model diagnostics, confidence intervals, and effect sizes alongside p-values, ensuring that stakeholders could interpret both statistical and practical significance and apply the results responsibly in policy, program design, and ongoing performance evaluation.

LIMITATIONS

This study had several limitations that should be considered when interpreting the results and applying the findings to other contexts. First, the quantitative evidence was derived from a single cross-sectional dataset, which limited causal interpretation because temporal ordering between predictors and the dependent variable could not be definitively established. Although regression modeling identified statistically significant associations and quantified unique predictive contributions, the results should be interpreted as explanatory and predictive relationships rather than proof of causation. Second, the study relied on self-reported responses for the constructs, which introduced the potential for common method variance, social desirability bias, and recall-related measurement error. Even when reliability coefficients indicated acceptable internal consistency, reliability did not fully eliminate the risk that respondents interpreted items differently or responded in patterned ways based on mood, perceived expectations, or survey fatigue. Third, the demographic composition of the sample may have influenced generalizability, particularly if participation was concentrated in specific educational, employment, or geographic segments. Overrepresentation of certain groups could have shaped construct means, variances, and regression coefficients, limiting the extent to which the findings could be transferred to populations with different characteristics. Fourth, the non-significant predictor in the multivariate model suggested that construct overlap, measurement sensitivity, or indirect effects may have affected interpretability. The study did not test mediation or moderation mechanisms, and therefore could not determine whether the unsupported relationship operated indirectly through other constructs or became relevant only under specific conditions. Fifth, although diagnostic indicators suggested that multicollinearity was not severe, moderate inter-construct correlations could still have attenuated certain coefficients by distributing shared variance across predictors, which may have reduced the observed direct effect for weaker constructs. Sixth, the analysis used a single primary modeling approach; alternative specifications, additional control variables, or non-linear models might have yielded different effect estimates or revealed interaction effects not captured in the linear framework. Finally, because the study focused on statistical significance and explained variance within the available data, external validity depended on replication with independent samples, longitudinal designs, and broader contextual settings. These limitations indicated that while the findings were statistically robust within the study sample, continued validation, model refinement, and context-specific testing remained necessary to strengthen generalizability and practical application.

REFERENCES

[1]. Allan, B. A., Batz-Barbarich, C., Sterling, H. M., & Tay, L. (2019). Outcomes of meaningful work: A meta-analysis. *Journal of management studies*, *56*(3), 500-528.

[2]. An, S., Malhotra, K., Dilley, C., Han-Burgess, E., Valdez, J. N., Robertson, J., Clark, C., Westover, M. B., & Sun, J. (2018). Predicting drug-resistant epilepsy—A machine learning approach based on administrative claims data. *Epilepsy & Behavior*, *89*, 118-125.

[3]. Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, *30*, 98-108.

[4]. Becker, D., van Breda, W., Funk, B., Hoogendoorn, M., Ruwaard, J., & Riper, H. (2018). Predictive modeling in e-mental health: a common language framework. *Internet interventions*, *12*, 57-67.

[5]. Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, *26*(1), 87-122.

[6]. Blasco, M. J., Castellví, P., Almenara, J., Lagares, C., Roca, M., Sesé, A., Piqueras, J. A., Soto-Sanz, V., Rodríguez-Marín, J., & Echeburúa, E. (2016). Predictive models for suicidal thoughts and behaviors among Spanish University students: rationale and methods of the UNIVERSAL (University & mental health) project. *BMC psychiatry*, *16*(1), 122.

[7]. Button, K. S., Kounali, D., Thomas, L., Wiles, N. J., Peters, T., Welton, N. J., Ades, A., & Lewis, G. (2015). Minimal clinically important difference on the Beck Depression Inventory-II according to the patient's perspective. *Psychological medicine*, *45*(15), 3269-3279.

[8]. Buus, N., Caspersen, J., Hansen, R., Stenager, E., & Fleischer, E. (2014). Experiences of parents whose sons or daughters have (had) attempted suicide. *Journal of Advanced Nursing*, *70*(4), 823-832.

[9]. Chen, D., Zhang, R., Zhao, H., & Feng, J. (2019). A Bibliometric Analysis of the Development of ICD-11 in Medical Informatics. *Journal of Healthcare Engineering*, *2019*(1), 1649363.

[10]. Cheng, H., Clymer, J. W., Chen, B. P.-H., Sadeghirad, B., Ferko, N. C., Cameron, C. G., & Hinoul, P. (2018). Prolonged operative duration is associated with complications: a systematic review and meta-analysis. *Journal of Surgical Research*, *229*, 134-144.

[11]. Chiu, C. M., Wang, E. T., Fang, Y. H., & Huang, H. Y. (2014). Understanding customers' repeat purchase intentions in B2C e-commerce: the roles of utilitarian value, hedonic value and perceived risk. *Information systems journal*, *24*(1), 85-114.

[12]. Clement, N., MacDonald, D., & Simpson, A. (2014). The minimal clinically important difference in the Oxford knee score and Short Form 12 score after total knee arthroplasty. *Knee Surgery, Sports Traumatology, Arthroscopy*, *22*(8), 1933-1939.

[13]. Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. *Journal of business and psychology*, *29*(1), 1-19.

[14]. Duwe, G., & Kim, K. (2016). Sacrificing accuracy for transparency in recidivism risk assessment: The impact of classification method on predictive performance. *Corrections*, *1*(3), 155-176.

[15]. Edgcomb, J., Shaddox, T., Hellemann, G., & Brooks III, J. O. (2019). High-risk phenotypes of early psychiatric readmission in bipolar disorder with comorbid medical illness. *Psychosomatics*, *60*(6), 563-573.

[16]. Feldman, S., & Johnston, C. (2014). Understanding the determinants of political ideology: Implications of structural complexity. *Political Psychology*, *35*(3), 337-358.

[17]. Gandhi, A., Luyckx, K., Maitra, S., & Claes, L. (2016). Non-suicidal self-injury and other self-directed violent behaviors in India: A review of definitions and research. *Asian journal of psychiatry*, *22*, 196-201.

[18]. Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and computing*, *24*(6), 997-1016.

[19]. Giner, L., Guija, J. A., Root, C. W., & Baca-Garcia, E. (2016). Nomenclature and definition of suicidal Behavior. In *Understanding suicide: From diagnosis to personalized treatment* (pp. 3-17). Springer.

[20]. Goodfellow, B., Kolves, K., & De Leo, D. (2019). Contemporary definitions of suicidal behavior: a systematic literature review. *Suicide and Life-Threatening Behavior*, *49*(2), 488-504.

[21]. Grunert, K. G., Hieke, S., & Wills, J. (2014). Sustainability labels on food products: Consumer motivation, understanding and use. *Food policy*, *44*, 177-189.

[22]. Habous, M., Tal, R., Tealab, A., Soliman, T., Nassar, M., Mekawi, Z., Mahmoud, S., Abdelwahab, O., Elkhouly, M., & Kamr, H. (2018). Defining a glycated haemoglobin (HbA1c) level that predicts increased risk of penile implant infection. *BJU international*, *121*(2), 293-300.

[23]. Haque, B. M. T., & Md. Arifur, R. (2021). ERP Modernization Outcomes in Cloud Migration: A Meta-Analysis of Performance and Total Cost of Ownership (TCO) Across Enterprise Implementations. *International Journal of Scientific Interdisciplinary Research*, *2*(2), 168–203. https://doi.org/10.63125/vrz8hw42

[24]. Hayes, A. F. (2018). Partial, conditional, and moderated moderated mediation: Quantification, inference, and interpretation. *Communication monographs*, *85*(1), 4-40.

[25]. Hone, K. S., & El Said, G. R. (2016). Exploring the factors affecting MOOC retention: A survey study. *Computers & Education*, *98*, 157-168.

[26]. Hruska, B., Irish, L. A., Pacella, M. L., Sledjeski, E. M., & Delahanty, D. L. (2014). PTSD symptom severity and psychiatric comorbidity in recent motor vehicle accident victims: A latent class analysis. *Journal of anxiety disorders*, *28*(7), 644-649.

[27]. Jakobsen, J. C., Wetterslev, J., Winkel, P., Lange, T., & Gluud, C. (2014). Thresholds for statistical and clinical significance in systematic reviews with meta-analytic methods. *BMC medical research methodology*, *14*(1), 120.
[28]. Jones, P., & Bowes, J. (2017). Rendering systems visible for design: Synthesis maps as constructivist design narratives. *She Ji: The Journal of Design, Economics, and Innovation*, *3*(3), 229-248.
[29]. Kawamura, T., Egami, S., Tamura, K., Hokazono, Y., Ugai, T., Koyanagi, Y., Nishino, F., Okajima, S., Murakami, K., & Takamatsu, K. (2019). Report on the first knowledge graph reasoning challenge 2018: Toward the eXplainable AI system. Joint International Semantic Technology Conference,
[30]. Kyratsis, Y., Ahmad, R., Iwami, M., Castro-Sánchez, E., Atun, R., & Holmes, A. H. (2019). A multilevel neo-institutional analysis of infection prevention and control in English hospitals: coerced safety culture change? *Sociology of health & illness*, *41*(6), 1138-1158.
[31]. Lee, E. C., Whitehead, A. L., Jacques, R. M., & Julious, S. A. (2014). The statistical interpretation of pilot trials: should significance thresholds be reconsidered? *BMC medical research methodology*, *14*(1), 41.
[32]. Lee, J., Jang, H., Kim, J., & Min, S. (2019). Development of a suicide index model in general adolescents using the South Korea 2012–2016 national representative survey data. *Scientific reports*, *9*(1), 1846.
[33]. Liem, C. C., Langer, M., Demetriou, A., Hiemstra, A. M., Sukma Wicaksana, A., Born, M. P., & König, C. J. (2018). Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening. In *Explainable and interpretable models in computer vision and machine learning* (pp. 197-253). Springer.
[34]. Lim, J., Rekhi, G., Rapisarda, A., Lam, M., Kraus, M., Keefe, R. S., & Lee, J. (2015). Impact of psychiatric comorbidity in individuals at Ultra High Risk of psychosis—Findings from the Longitudinal Youth at Risk Study (LYRIKS). *Schizophrenia research*, *164*(1-3), 8-14.
[35]. Liu, C., Ishi, C., & Ishiguro, H. (2019). Probabilistic nod generation model based on speech and estimated utterance categories. *Advanced Robotics*, *33*(15-16), 731-741.
[36]. Maier-Hein, L., Vedula, S. S., Speidel, S., Navab, N., Kikinis, R., Park, A., Eisenmann, M., Feussner, H., Forestier, G., & Giannarou, S. (2017). Surgical data science for next-generation interventions. *Nature Biomedical Engineering*, *1*(9), 691-696.
[37]. McGinnis, R. S., McGinnis, E. W., Hruschak, J., Lopez-Duran, N. L., Fitzgerald, K., Rosenblum, K. L., & Muzik, M. (2018). Rapid anxiety and depression diagnosis in young children enabled by wearable sensors and machine learning. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),
[38]. McGlashan, A., Verrinder, G., & Verhagen, E. (2018). Working towards more effective implementation, dissemination and scale-up of lower-limb Injury-Prevention programs: insights from community Australian football coaches. *International journal of environmental research and public health*, *15*(2), 351.
[39]. Md Ashraful, A., Md Fokhrul, A., & Md Fardaus, A. (2020). Predictive Data-Driven Models Leveraging Healthcare Big Data for Early Intervention And Long-Term Chronic Disease Management To Strengthen U.S. National Health Infrastructure. *American Journal of Interdisciplinary Studies*, *1*(04), 26-54. https://doi.org/10.63125/1z7b5v06
[40]. Md Fokhrul, A., Md Ashraful, A., & Md Fardaus, A. (2021). Privacy-Preserving Security Model for Early Cancer Diagnosis, Population-Level Epidemiology, And Secure Integration into U.S. Healthcare Systems. *American Journal of Scholarly Research and Innovation*, *1*(02), 01–27. https://doi.org/10.63125/q8wjee18
[41]. Metzger, M. H., Tvardik, N., Gicquel, Q., Bouvry, C., Poulet, E., & Potinet-Pagliaroli, V. (2017). Use of emergency department electronic medical records for automated epidemiological surveillance of suicide attempts: a French pilot study. *International journal of methods in psychiatric research*, *26*(2), e1522.
[42]. Miklin, S., Mueller, A. S., Abrutyn, S., & Ordonez, K. (2019). What does it mean to be exposed to suicide?: Suicide exposure, suicide risk, and the importance of meaning-making. *Social Science & Medicine*, *233*, 21-27.
[43]. Mouheb, D., Albarghash, R., Mowakeh, M. F., Al Aghbari, Z., & Kamel, I. (2019). Detection of Arabic cyberbullying on social networks using machine learning. 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA),
[44]. Nguyen, T., O'Dea, B., Larsen, M., Phung, D., Venkatesh, S., & Christensen, H. (2017). Using linguistic and topic analysis to classify sub-groups of online depression communities. *Multimedia tools and applications*, *76*(8), 10653-10676.
[45]. Ngwena, J., Hosany, Z., & Sibindi, I. (2017). Suicide: a concept analysis. *Journal of Public Health*, *25*(2), 123-134.
[46]. Olsen, M. F., Bjerre, E., Hansen, M. D., Hilden, J., Landler, N. E., Tendal, B., & Hróbjartsson, A. (2017). Pain relief that matters to patients: systematic review of empirical studies assessing the minimum clinically important difference in acute pain. *BMC medicine*, *15*(1), 35.
[47]. Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language learning*, *64*(4), 878-912.
[48]. Price, W., & Nicholson, I. (2019). Medical AI and contextual bias. *Harv. JL & Tech.*, *33*, 65.
[49]. Rankin, J., Matthews, L., Cobley, S., Han, A., Sanders, R., Wiltshire, H. D., & Baker, J. S. (2016). Psychological consequences of childhood obesity: psychiatric comorbidity and prevention. *Adolescent health, medicine and therapeutics*, 125-146.
[50]. Rauf, M. A. (2018). A needs assessment approach to english for specific purposes (ESP) based syllabus design in Bangladesh vocational and technical education (BVTE). *International Journal of Educational Best Practices*, *2*(2), 18-25.
[51]. Rimkeviciene, J., & De Leo, D. (2015). Impulsive suicide attempts: a systematic literature review of definitions, characteristics and risk factors. *Journal of affective disorders*, *171*, 93-104.
[52]. Rolfson, O., Bohm, E., Franklin, P., Lyman, S., Denissen, G., Dawson, J., Dunn, J., Eresian Chenok, K., Dunbar, M., & Overgaard, S. (2016). Patient-reported outcome measures in arthroplasty registries: report of the Patient-

Reported Outcome Measures Working Group of the International Society of Arthroplasty Registries Part II. Recommendations for selection, administration, and analysis. *Acta orthopaedica*, *87*(sup1), 9-23.

[53]. Rosic, T., Naji, L., Bawor, M., Dennis, B. B., Plater, C., Marsh, D. C., Thabane, L., & Samaan, Z. (2017). The impact of comorbid psychiatric disorders on methadone maintenance treatment in opioid use disorder: a prospective cohort study. *Neuropsychiatric disease and treatment*, 1399-1408.

[54]. Saha, B., Nguyen, T., Phung, D., & Venkatesh, S. (2016). A framework for classifying online mental health-related communities with an interest in depression. *IEEE journal of biomedical and health informatics*, *20*(4), 1008-1015.

[55]. Salgado, R., Denkert, C., Demaria, S., Sirtaine, N., Klauschen, F., Pruneri, G., Wienert, S., Van den Eynden, G., Baehner, F. L., & Pénault-Llorca, F. (2015). The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Annals of oncology*, *26*(2), 259-271.

[56]. Sanderson, M., Bulloch, A. G., Wang, J., Williamson, T., & Patten, S. B. (2019). Predicting death by suicide using administrative health care system data: can feedforward neural network models improve upon logistic regression models? *Journal of affective disorders*, *257*, 741-747.

[57]. Sau, A., & Bhakta, I. (2019). Screening of anxiety and depression among seafarers using machine learning technology. *Informatics in Medicine Unlocked*, *16*, 100228.

[58]. Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental science*, *20*(3), e12372.

[59]. Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham L. Rev.*, *87*, 1085.

[60]. Sharif, W., Mumtaz, S., Shafiq, Z., Riaz, O., Ali, T., Husnain, M., & Choi, G. S. (2019). An empirical approach for extreme behavior identification through tweets using machine learning. *Applied Sciences*, *9*(18), 3723.

[61]. Shoib, S., & Kim, Y.-K. (2019). The frontiers of suicide. In *Frontiers in Psychiatry: Artificial Intelligence, Precision Medicine, and Other Paradigm Shifts* (pp. 503-517). Springer.

[62]. Silverman, M. M. (2016). Challenges to defining and classifying suicide and suicidal behaviors. *The international handbook of suicide prevention*, 9-35.

[63]. Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of suicide ideation in social media forums using deep learning. *Algorithms*, *13*(1), 7.

[64]. Talati, A., Keyes, K., & Hasin, D. (2016). Changing relationships between smoking and psychiatric disorders across twentieth century birth cohorts: clinical and research implications. *Molecular psychiatry*, *21*(4), 464-471.

[65]. Thaipisuttikul, P., Ittasakul, P., Waleeprakhon, P., Wisajun, P., & Jullagate, S. (2014). Psychiatric comorbidities in patients with major depressive disorder. *Neuropsychiatric disease and treatment*, 2097-2103.

[66]. Thompson, W. K., Gershon, A., O'Hara, R., Bernert, R. A., & Depp, C. A. (2014). The prediction of study-emergent suicidal ideation in bipolar disorder: a pilot study using ecological momentary assessment data. *Bipolar Disorders*, *16*(7), 669-677.

[67]. Torii, M., Fan, J.-w., Yang, W.-l., Lee, T., Wiley, M. T., Zisook, D. S., & Huang, Y. (2015). Risk factor detection for heart disease by applying text analytics in electronic medical records. *Journal of biomedical informatics*, *58*, S164-S170.

[68]. Tymofiyeva, O., Yuan, J. P., Huang, C.-Y., Connolly, C. G., Blom, E. H., Xu, D., & Yang, T. T. (2019). Application of machine learning to structural connectome to predict symptom reduction in depressed adolescents with cognitive behavioral therapy (CBT). *NeuroImage: Clinical*, *23*, 101914.

[69]. Valerio, M., Donaldson, I., Emberton, M., Ehdaie, B., Hadaschik, B. A., Marks, L. S., Mozer, P., Rastinehad, A. R., & Ahmed, H. U. (2015). Detection of clinically significant prostate cancer using magnetic resonance imaging–ultrasound fusion targeted biopsy: a systematic review. *European urology*, *68*(1), 8-19.

[70]. Van Ryn, M., Hardeman, R., Phelan, S. M., Dovidio, J. F., Herrin, J., Burke, S. E., Nelson, D. B., Perry, S., Yeazel, M., & Przedworski, J. M. (2015). Medical school experiences associated with change in implicit racial bias among 3547 students: a medical student CHANGES study report. *Journal of general internal medicine*, *30*(12), 1748-1756.

[71]. Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2018). Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. *Journal of child psychology and psychiatry*, *59*(12), 1261-1270.

[72]. Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry*, *14*(3), 270-277.

[73]. Weaver, C. M., Gordon, C. M., Janz, K. F., Kalkwarf, H., Lappe, J. M., Lewis, R., O'Karma, M., Wallace, T. C., & Zemel, B. (2016). The National Osteoporosis Foundation's position statement on peak bone mass development and lifestyle factors: a systematic review and implementation recommendations. *Osteoporosis international*, *27*(4), 1281-1386.

[74]. Yarnell, L. M., Stafford, R. E., Neff, K. D., Reilly, E. D., Knox, M. C., & Mullarkey, M. (2015). Meta-analysis of gender differences in self-compassion. *Self and identity*, *14*(5), 499-520.

[75]. Yazhini, K., & Loganathan, D. (2019). A state of art approaches on deep learning models in healthcare: An application perspective. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI),

[76]. Zaman, M. A. U., Sultana, S., Raju, V., & Rauf, M. A. (2021). Factors İmpacting the Uptake of İnnovative Open and Distance Learning (ODL) Programmes in Teacher Education. *Turkish Online Journal of Qualitative Inquiry*, *12*(6).

[77]. Zolotas, M., Elsdon, J., & Demiris, Y. (2018). Head-mounted augmented reality for explainable robotic wheelchair assistance. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),