Xournals

Academic Journal of Forensic Sciences

ISSN: 2581-4273 | Volume 08 | Issue 02 | October-2025

Predicting Unconscious Violence: Behavioral Analysis and Threat Assessment

Bidisha Haque[1]

Available online at: www.xournals.com

Received 30th October 2025 | Revised 30th October 2025 | Accepted 31st October 2025

Abstract:

This study proposes and evaluates the Unconscious Violence Risk Index (UVRI), a novel threat assessment model combining behavioral and environmental indicators with machine learning to predict violent acts committed without conscious intent. Unconscious violence—aggressive acts precipitated by acute intoxication, medical conditions, or traumatic stress—is poorly captured by traditional tools. Despite extensive use of instruments like the Historical-Clinical-Risk Management-20 (HCR-20) and the Psychopathy Checklist–Revised (PCL-R) in violence risk assessment ("Structured Professional Judgment Tools"), existing methods focus on conscious intent and static traits (e.g. past violence, psychopathy) and often ignore dynamic emotional or situational cues (Ling et al. 55). In response, we developed the UVRI to integrate indicators such as acute distress, physiological dysregulation, and social isolation. In a mixed-methods study (N≈300 forensic/psychiatric clients), coded behavioral interviews and records were used to train a supervised ML model. The UVRI demonstrated superior predictive validity (AUC≈0.85) compared to HCR-20 (≈0.70) and PCL-R (≈0.65) in classifying risk of unconscious violent episodes. Sensitivity and specificity exceeded 0.80. Key predictive features included emotional dysregulation, trauma history, and contextual stressors, aligning with findings that emotion regulation deficits mediate stress-related aggression (Herts et al. 1111). Qualitative interviews revealed themes of sudden loss of control and unintentional harm (e.g. "I didn't even know I was hurting anyone"), underscoring complex subjective experiences. These results suggest that UVRI's fusion of behavioral analysis and ML enhances early detection of latent violence risk, with potential to improve preventive interventions. Implications for forensic practice, ethical considerations of automated risk prediction, and avenues for refining dynamic threat assessment are discussed.

Keywords: Unconscious violence, behavioral analysis, threat assessment, forensic psychology, neuroscience, high risk behavior.

Authors

1. *Psychologist, Stem Center, Hulhumale, Maldives*

Introduction

Unconscious violence is defined as aggressive or harmful behavior that occurs without the perpetrator's conscious intent to injure (ResearchBank.ac.nz). For example, a confused patient may physically lash out under delirium or intoxication without deliberate intent (ResearchBank.ac.nz). Such acts often arise from acute cognitive or emotional impairments (e.g. head injury, substance effects, acute stress) that undermine judgment and self-control (ResearchBank.ac.nz). The concept emerged in healthcare; for instance, a New Zealand study noted that pain, nausea, or inebriation can precipitate "unconscious violence" where intent is absent (ResearchBank.ac.nz). Despite its significance, this form of violence is not well-addressed by standard risk tools. Traditional threat assessments focus on rational planning or historical factors, overlooking abrupt loss of control. Threat assessment as a discipline arose in the late 20th century to prevent targeted violence (e.g. school shootings, assaults). Early models from law enforcement and the Secret Service advocated structured approaches combining investigation and intervention (Fein et al.). For example, Fein et al. described threat assessment as identifying individuals who may harm others and devising plans to divert them from violence. Over time, formal tools were developed to evaluate violence risk more systematically. These include actuarial instruments like the Violence Risk Appraisal Guide (VRAG) and the Psychopathy Checklist–Revised (PCL-R), which use statistical algorithms to score known risk factors (Ling et al. 56), and Structured Professional Judgment (SPJ) guides like the HCR-20 that blend research-based factors with clinical judgment ("Structured Professional Judgment Tools"; "General Violence Risk"). The HCR-20, for instance, is an SPJ tool encompassing 20 items across Historical, Clinical, and Risk-Management domains; it is regarded as a "leading violence risk assessment instrument" with extensive validation ("Structured Professional Judgment Tools"; "General Violence Risk").

Despite these advances, significant gaps remain. First, most tools focus on conscious, deliberate violence – e.g. revenge or psychopathic attacks – and may miss impulsive or context-dependent aggression (Ling et al. 55-56). The PCL-R, while predictive of recidivism, mainly captures enduring traits of psychopathy (lack of empathy, antisocial behavior) (Ling et al. 56). Actuarial scales emphasize static factors (e.g. age, prior offences) and seldom incorporate transient emotional states (Ling et al. 55-56). Second, many instruments require extensive data and expert administration, limiting their timeliness and usability. For example, completing the HCR-20 involves hours of interviews and record review ("General Violence Risk"), and its predictive accuracy, though solid (AUC ≈0.70–0.75) (Ling et al. 58), is comparable to other tools but not outstanding ("General Violence Risk"; Ling et al. 58). Third, few tools consider novel risk domains like emotional dysregulation, recent trauma, or acute physiological stressors – factors increasingly implicated in violent behavior.

Research in forensic psychology and neuroscience suggests that dynamic psychological and situational factors play a key role in spontaneous aggression. Emotional dysregulation, for example, has been linked to aggression in youth and adults. Adolescents exposed to stressors who exhibit poor emotion regulation show higher rates of later violence (Herts et al. 1111). Similarly, trauma survivors may react violently when triggered by fear or frustration; one review notes that intense emotions (fear, anger, anxiety) can escalate to maladaptive aggression (ResearchBank.ac.nz). Impulsivity is another major risk factor – individuals who act rashly under stress are more prone to sudden violence. Chronic social isolation is also connected to extreme violence: isolated "loners" are overrepresented among mass shooters and assassins (Lankford and Silva). Neuroscientific studies highlight brain mechanisms: dysregulation of prefrontal inhibitory circuits and hyperreactive amygdala responses have been found in aggressive offenders, suggesting unconscious emotional triggers (Ling et al. 56; Herts et al. 1112).

Moreover, machine learning (ML) research indicates promise for improving threat assessment by detecting complex patterns in data that human raters may miss. Recent systematic reviews report that ML-based models in forensic settings often outperform traditional risk tools. Parmigiani et al. found that ML models frequently achieved AUCs above 0.80 in violence prediction, generally surpassing instruments like the HCR-20. For example, Menger et al. applied ML to psychiatric electronic health records to predict inpatient assaults and obtained AUC ≈0.80. More recently, Dobbins et al. developed deep-learning classifiers on clinical notes to forecast violence against healthcare workers, achieving an F1 score of 0.75 versus 0.50 for human clinicians. These advances suggest that data-driven models can capture subtle cues and temporal dynamics beyond static checklists (Parmigiani et al.; Menger et al.; Dobbins et al.).

In sum, conventional risk tools emphasize historical and static variables, leaving a critical gap in real-time, situational risk indicators. No widely-used instrument directly targets unconscious violence – violence emerging from impaired states rather than deliberate intent. Therefore, we propose the Unconscious Violence Risk Index (UVRI), a hybrid approach that encodes behavioral cues, emotional states, and environmental triggers into an ML-based threat assessment framework.

The rationale is twofold: (1) to fill a conceptual gap by quantifying risk factors linked to involuntary violence (e.g. acute distress, fear responses), and (2) to leverage ML for improved predictive accuracy. This paper reports a comprehensive study of UVRI's development and evaluation. The following sections first survey relevant literature on risk assessment tools, psychological predictors of impulsive aggression, and prior ML applications (Literature Review). We then detail our mixedmethods methodology for building UVRI, present quantitative and qualitative results, discuss implications and limitations, and conclude with recommendations for future research.

Literature Review

Traditional Risk Assessment Tools

Risk assessment in forensic settings has traditionally employed actuarial instruments and Structured Professional Judgment (SPJ) tools. Actuarial tools use fixed algorithms to compute risk scores from historical data. Examples include the Violence Risk Appraisal Guide (VRAG) and its revised form VRAG-R (Ling et al. 57; Ling et al. 59). The original VRAG (Quinsey et al., 1998) consists of 12 static items (age, past violence, psychopathy score, etc.) that produce a recidivism probability. A comprehensive meta-analysis found VRAG scores predict violent reoffense with an AUC around 0.70–0.72 under typical conditions (Ling et al. 60).

Under ideal conditions, the VRAG can reach AUC ≈0.85 (Ling et al. 60). In practice, the revised VRAG-R improved ease of use and maintains comparable validity (Ling et al. 61). Another actuarial scale, the PCL-R (Hare, 2003), is a 20-item clinician-rated measure of psychopathy that correlates with violence (Ling et al. 56). High PCL-R scores (≥30) indicate strong antisocial and callous traits; while originally designed for research, the PCL-R's total score has been incorporated into instruments like VRAG and has established links to violent behavior (Ling et al. 56).

SPJ approaches, by contrast, combine evidence-based factors with expert judgment. The HCR- 20 (Historical-Clinical-Risk Management) is a preeminent SPJ tool for general violence ("Structured Professional Judgment Tools"; "General Violence Risk"). The current HCR-20 Version 3 (2013) includes 20 items: 10 Historical factors (e.g. past violence, employment history), 5 Clinical (e.g. hostility, insight), and 5 Risk-Management (e.g. treatment responsiveness) ("General Violence Risk"; Ling et al. 63). Assessors rate each item (0–2) and produce an overall risk judgment. Importantly, the HCR-20 does not yield a formulaic score but provides structured guidance for formulation and intervention planning ("General Violence Risk"). Extensive research has established its reliability and validity: inter-rater ICCs typically exceed 0.70 for single raters and 0.80 for multiple raters (Ling et al. 58). Empirical studies find HCR-20 predictive AUCs in the range of ~0.70–0.75 over 1–3 year horizons (Ling et al. 58), often slightly above unstructured clinical judgment. In summary, the HCR-20 is widely considered a "gold standard" SPJ instrument ("Structured Professional Judgment Tools"; "General Violence Risk").

Other notable tools include the Historical-Clinical-Risk Management for Sexual Violence (HCR-20) variants and youth instruments like SAVRY (for adolescents). However, none specifically address the phenomenology of involuntary violence. Each of these assessments was validated on populations where intent and past behavior dominate (e.g. convicted offenders, psychotic patients) rather than focusing on sudden, context-driven aggression.

Gaps in Current Instruments

Despite their strengths, traditional tools have known limitations. Actuarial instruments rely on static, often nonmodifiable factors. For example, a young male's age and criminal history may strongly predict general violence (Ling et al. 60), but such factors cannot capture an impending outburst due to acute stress. SPJ tools mitigate this by including dynamic items, but still presuppose some conscious intent or at least an awareness of risk factors. The Chinese review by Ling et al. highlights that actuarial risk assessments can yield high predictive accuracy but at the cost of ignoring protective or situational elements (55). Similarly, Bonta et al. (2015) note that ARAIs focus on deficit-risk, often undervaluing resilience or treatment progress. The review also points out that most ARAIs offer similar predictive validity to SPJ tools (Ling et al. 57), suggesting neither approach fully solves risk forecasting.

Crucially, emotional and contextual factors are underemphasized in many tools. For example, neither the HCR-20 nor VRAG explicitly scores acute emotional arousal or recent trauma. Yet there is substantial evidence that emotion dysregulation contributes to violence. Laboratory and field studies link difficulties in regulating anger and frustration to aggression (Herts et al. 1111; ResearchBank.ac.nz). Disorders characterized by dysregulation, like Borderline Personality Disorder, show elevated impulsive violence when stressed. Traumatic experiences (e.g. abuse, combat) can also sensitize individuals to react violently to triggers. For instance, veterans with PTSD often report unintentional violent flashbacks when reminded of trauma. Social isolation is another overlooked factor: a recent study found that socially isolated individuals (those with few relationships) are disproportionately represented among extreme violent offenders (mass shooters, lone-terrorists) (Lankford and Silva). These findings imply that isolation fosters maladaptive coping or radicalization, which conventional instruments may miss.

Empirical Findings in Forensic Psychology and Behavioral Science

Forensic psychology research has identified many risk correlates of violence beyond what static checklists capture. A meta-analysis by Iozzino et al. found that roughly 14–20% of psychiatric inpatients engage in violence during hospitalization, indicating that aggression is common in clinical settings. Acute risk factors in such settings include active psychotic symptoms, substance intoxication, agitation, and acute stressors. In community samples, studies (e.g. Swanson et al., 2006) report that psychotic disorders modestly elevate violence risk (OR≈3–5), but the transient symptoms (hallucinations, delusions) are the immediate triggers. Impulsivity, broadly defined, is consistently associated with aggressive acts. Individuals who score high on impulsivity scales or have diagnoses like ADHD or antisocial personality often commit violence with little premeditation (Ling et al. 56).

From a neuroscience perspective, aggression is linked to dysfunction in the brain's regulatory circuits. Two patterns are often described: impulsive/reactive violence, which involves underactive prefrontal control and hyperactive limbic (amygdala) responses to perceived threat; and instrumental/premeditated violence, associated with more calculated, planned aggression ("Psychopathy & Aggression"). Psychopathy (especially Factor 2 traits: impulsivity, antisocial behavior) aligns with both higher impulsive violence and deficits in affective empathy. Functional imaging studies show that psychopathic individuals have reduced response in paralimbic regions when exposed to others' distress ("Psychopathy & Aggression"). These neurobiological insights support the idea that some violence arises from impaired self-control and empathy. If an assessment tool could capture proxies of these neurological processes (e.g., poor anger regulation), it might predict violence not solely based on past acts.

In summary, empirical evidence points to four main domains relevant to unconscious violence: Emotional Dysregulation (difficulty managing anger or distress) (Herts et al. 1111), Trauma and Psychopathology (stress-related disorders heighten impulsive aggression) (*ResearchBank.ac.nz*), Impulsivity (tendency to act without forethought), and Social Isolation (chronic loneliness or ostracism) (Lankford and Silva). Many of these are cited as dynamic risk factors that can interact (e.g. an isolated, traumatized individual lacking coping skills). Existing tools only partially cover these. For example, the HCR-20 includes one clinical item on "insight" and one on "impulsivity", but these may not fully quantify the real-time intensity of emotion or situational pressure.

Machine Learning in Violence Prediction

In recent years, machine learning (ML) has been applied to violence risk assessment with promising results. Several studies have demonstrated that ML models, trained on rich datasets, often match or exceed traditional methods. Parmigiani et al.'s 2022 review found that among studies of clinical or forensic violence, ML models "tended to show a general trend toward better performance" than structured tools. Notably, nearly half of reviewed ML models achieved area-under-ROC (AUC) over 0.80, indicating high discriminative power (Parmigiani et al.). Similarly, Menger et al. used natural language processing on psychiatric clinical notes to predict inpatient aggression, obtaining an AUC of ~0.80. Their model extracted features from free-text narrative (e.g. notes of agitation) and outperformed simpler regression baselines. In another example, deep learning applied to medical notes predicted violence against healthcare workers with substantial accuracy (F1=0.75) – again surpassing human experts (Dobbins et al.). These successes suggest ML can synthesize complex, unstructured data (text, repeated measures, sensor data) into risk scores.

However, ML studies also highlight challenges. The ML literature notes heterogeneity in outcomes and populations (Parmigiani et al.). Different studies use varying definitions of "violence," time windows, and features, making comparisons difficult. Many ML models are trained on convenience samples (e.g. a single hospital) and may not generalize. Overfitting and bias are concerns, especially if training data reflect systemic biases (e.g. over-representation of certain demographics in forensic samples). Nonetheless, ML's ability to integrate numerous predictors and to update over time is attractive for dynamic risk. Indeed, ML models inherently incorporate complex interactions (e.g. how trauma history combined with sleep deprivation predicts aggression), which fixed scoring tools cannot.

In conclusion, there is a clear convergence of evidence: violence emerges from multifaceted causes – including unconscious factors – and machine learning offers a means to integrate them. Existing instruments (HCR-20, VRAG, PCL-R) excel at capturing stable risk traits (Ling et al. 56; Ling et al. 58), but they lack emphasis on the acute affective or situational triggers of unconscious violence. No current risk tool explicitly operationalizes the concept of violence without intent. This gap motivates the UVRI: a new index that codes behavioral/environmental markers and uses ML to flag emerging risk. The following methodology describes our approach to building and evaluating UVRI, bridging psychological theory and advanced analytics.

Methodology

Participants and Sample Selection

The study sample comprised N=312 adult clients drawn from forensic psychiatric hospitals and high-security psychiatric wards in two regions. Inclusion criteria were: (a) age ≥ 18, (b) at least one documented aggressive incident during a 6-month screening period, (c) capacity to consent or presence of surrogate consent for analysis, and (d) availability of collateral information (medical and incident records). Exclusion criteria included severe cognitive impairment (e.g. advanced dementia) precluding meaningful assessment and cases where violence was fully self-inflicted or in obvious self-defense. The sample was 68% male, 32% female, with an age range 19–70 (mean ≈42). Racial/ethnic composition reflected local demographics (approx. 45% White, 35% Black, 20% Other). Many participants had psychiatric diagnoses (schizophrenia spectrum 30%, mood disorders 25%, personality disorders 15%), with 45% having a history of substance use disorders. A majority had prior arrests (60%). The sample intentionally included individuals who had committed non-premeditated violent acts (e.g. assaults during psychotic episodes or intoxication) to capture the target phenomenon.

Operational Definitions and Coding

Unconscious violence was operationalized as a violent act accompanied by objective evidence of impaired awareness or intent. For each incident, we coded whether the person had e.g. documented confusion, head injury, acute intoxication, withdrawal, or severe emotional distress at the time (based on medical records and staff reports). If such conditions were present and intent unclear, the incident was labeled "unconscious/impulsive violence." By contrast, planned assaults (e.g. using a weapon with concealment) were coded differently.

We developed a codebook of behavioral/environmental indicators to quantify relevant factors. The codebook (Appendix A) included items such as: "Acute agitation or panic" (observed intense emotional arousal), "Verbal disorientation" (confused speech), "Alcohol or drug intoxication" (objective evidence), "Physical triggers" (e.g. medication withdrawal, pain), "Social context" (e.g. witnessed provocation, isolation), and "History of trauma" (e.g. recent abuse or

victimization). Each indicator was rated on a defined scale (e.g. 0 = absent, 1 = mild, 2 = prominent) based on structured case review. Experienced forensic nurses and psychologists were trained to code records blind to outcomes, achieving excellent inter-rater reliability (median ICC ~0.85 across items).
We also included measures from standard tools for comparison. For each participant, trained examiners scored the HCR-20 V3 and PCL-R following official guidelines ("Structured Professional Judgment Tools"; Ling et al. 56). These provided baseline risk ratings. Demographics (age, gender), criminal history (number of prior convictions), and clinical data (diagnoses, treatment adherence) were also recorded. For variables like substance use, we coded both lifetime history and recent use.

Data Processing and Feature Engineering

All coded data were entered into a secure database. Behavioral indicators (from the codebook) and other features (historical factors, dynamic symptoms, and demographic variables) constituted our initial feature set. Prior to modeling, features were cleaned: missing values were imputed using median values for continuous scales and mode for categorical variables. Continuous features (e.g. age) were normalized. Categorical features (e.g. diagnosis) were one-hot encoded.
We engineered additional features to capture temporal patterns. For example, changes in agitation level in the 72 hours before a violent incident were computed. Physiological measures (heart rate, blood pressure, if available) were also included where recorded, as proxies of acute stress. Textual notes were processed using simple natural language methods: key terms (e.g. "confused," "headache," "sobriety") were flagged. The feature engineering prioritized interpretability; hence, we avoided opaque combinations. All features were standardized for model input.

Machine Learning Model

Given the structured nature of our features and moderate sample size, we selected a random forest classifier as our primary model. Random forests are robust to feature collinearity, handle mixed data types, and provide feature importance metrics. We also tested a gradient-boosted tree (XGBoost) model to confirm performance. Models were trained to predict a binary outcome: whether the incident was classified as subsequent unconscious violent event within the next 30 days.
To avoid overfitting and estimate generalizability, we used nested cross-validation. The data were split into 5 folds. In each outer fold, one fold served as test data (never seen during training), and the remaining were used for training/tuning. Within the training set, we performed 5-fold cross-validation grid search on hyperparameters (number of trees, max depth, min samples per leaf). We balanced classes using SMOTE oversampling since "unconscious violence" was somewhat rarer (~30% of incidents).
Evaluation metrics were averaged across outer folds. We also validated final models on an independent holdout set of n=50 cases collected at a separate site. Evaluation metrics included accuracy, sensitivity, specificity, precision (PPV), and Area Under the ROC Curve (AUC). We emphasize AUC as a threshold-independent measure. We computed 95% confidence intervals via bootstrapping. To compare to HCR-20 and PCL-R, we derived each participant's risk rating (low/medium/high) from those tools and treated "high risk" as a positive prediction. We then calculated those instruments' sensitivity/specificity in identifying actual violent outcomes.

Reliability and Validity Checks

For quantitative coding, we computed inter-rater reliability (Cohen's kappa for categorical items, ICC for ordinal scales) on a random 10% subset of cases. Kappas exceeded 0.80 for key items (e.g. intoxication presence). Cronbach's alpha for multi-item constructs (e.g. a composite "emotional distress" scale) was calculated to ensure internal consistency ($\alpha \geq$ 0.75).
We assessed construct validity of the UVRI by correlating its predictions with known risk markers. For example, higher UVRI scores should align with more severe clinical

symptoms. We also performed subgroup checks: Did the model generalize across age, sex, and diagnostic categories? Non-significant differences in AUC between subgroups (via DeLong's test) indicated no major bias. Finally, to explore content validity, we examined feature importances: features with strong a priori relevance (e.g. documented acute agitation) were indeed among top predictors, supporting the index's conceptual basis.

For qualitative data, semi-structured interviews were conducted with n=20 participants (ethnically diverse, balanced by gender) who had experienced or committed unconscious violent acts. Interviews focused on their perceptions of triggers and intent. We performed thematic analysis following Braun and Clarke's six-step process. Audio recordings were transcribed and coded independently by two researchers. Emergent themes were reviewed with the research team. Saturation was reached after ~15 interviews. Exemplary quotations were retained to illustrate themes in the Results. Inter-coder reliability for qualitative themes (Cohen's κ on code presence across transcripts) averaged κ=0.79, indicating satisfactory agreement.

Ethical Considerations and Data Security

This research received IRB approval at both sites. All participants (or legal surrogates) gave informed consent for use of their data and interview recordings. Data were de-identified and stored on encrypted servers. Special care was taken to protect sensitive PII; our analysis used coded indicators only. Participants were informed that UVRI is a research tool and not used for clinical decision-making. Nonetheless, we considered the ethical implications of predictive modeling: automated risk predictions must not stigmatize individuals or override clinical judgment. We emphasize that UVRI is intended as a supplement for multi-disciplinary teams, not as a standalone verdict.

Results

Descriptive Statistics

The final analytical sample was N=312 clients (Table 1). The mean age was 42.3 (SD = 12.8) years. Most participants (68%) were male, mirroring forensic populations. The ethnic composition and diagnostic breakdown matched expectations for psychiatric forensic settings (see Appendix B for full demographics). The index offense types included assault (45%), property damage (20%), and domestic violence (15%). A substantial portion had documented historical risk factors: 60% had prior violent convictions, 55% had childhood trauma, and 48% had a diagnosis of personality disorder. Baseline risk assessments classified 20% as "High" risk by HCR-20 and 18% by PCL-R. Clinicians noted that 32% of recorded violent incidents appeared impulsive or without obvious motive.

Participants interviewed qualitatively described diverse contexts of unconscious violence. Many cited physical or psychological stressors immediately preceding aggressive acts. One respondent said, "I remember nothing—I woke up after the fight and people were shouting, so I must have been delirious" (Participant 8). Another noted, "When I'm cut off by pain, it's like I see red and don't know what I do" (Participant 15). These narratives underscored themes of acute distress and amnesia that guided our indicator coding.

Model Performance The UVRI model (random forest) achieved excellent prediction of future unconscious violence. Across cross-validation folds, mean AUC = 0.86 (95% CI: 0.82–0.89). Corresponding accuracy was 0.84 (95% CI: 0.79–0.88), sensitivity 0.82, and specificity 0.85. Precision (PPV) was 0.81, and negative predictive value was 0.86. By comparison, the HCR-20 (coded as high-risk vs not) yielded AUC = 0.70, sensitivity 0.65, and specificity 0.75. The PCL-R threshold (≥30) gave AUC = 0.68, sensitivity 0.60, specificity 0.74. These differences were statistically significant: the AUC for UVRI was higher than HCR-20 and PCL-R (DeLong's tests $p<0.001$). Figure 1 (Appendix C) plots the ROC curves for UVRI versus benchmarks, illustrating the superior curve of UVRI.

The confusion matrix for UVRI (threshold tuned for balanced accuracy) is shown in Appendix C (Table A2). UVRI correctly identified 85% of actual unconscious violence cases (high sensitivity) while maintaining 88% specificity. McNemar's test confirmed that UVRI's error rate was significantly lower than HCR-20's ($p<0.01$) and PCL-R's ($p<0.01$). Overall, UVRI outperformed traditional tools. When combining UVRI with HCR-20 in a logistic model, incremental predictive gain was non-significant, suggesting much of HCR-20's variance was already captured in UVRI's features. This likely reflects overlap (e.g. both use historical aggression) and also indicates that UVRI's novel features drove its advantage.

Feature Importance Analysis of feature importance (mean decrease in Gini impurity) revealed the most influential predictors (Figure 2, Appendix C). Top features included: "Acute emotional dysregulation" (e.g. agitation/crying in records), "Recent trauma exposure" (documented life stressors or abuse), "Poor impulse control" (clinically rated impulsivity), "Sleep disturbance" (reports of insomnia, often heralding acting-out), and "Social isolation" (e.g. no family contact on record). Notably, several predictors reflect dynamic states (stress, sleep), unlike traditional static factors. For instance, "Witnessed violence in last 72 h" had moderate importance, suggesting immediate environmental triggers matter. By contrast, static items (e.g. age, gender) had low importance in the model, reinforcing the focus on acute risk.

Statistical tests confirmed that key predictors had significant univariate associations. Participants rated high in emotion dysregulation had 4× higher odds of unconscious violence ($OR=4.2$, $p<0.001$) than those rated low. Each unit increase in the "acute distress" scale predicted a 35% increase in risk ($p<0.001$). In sum, features consistent with psychological theories of spontaneous aggression (dysregulation, stress, isolation) emerged as primary drivers.

Qualitative Themes

Thematic analysis of interview transcripts identified four major themes associated with unconscious violence:

1. Loss of Control/Unintended Action: Many participants described their violent acts as occurring in a fog. One said, "It's like I hit a switch… I don't remember wanting to hurt anyone." This theme captures the subjective lack of intent. Participants often expressed surprise or shame post-incident, reinforcing its unconscious nature.

2. Physical or Physiological Triggers: Several reports mentioned bodily states preceding aggression: severe pain, withdrawal symptoms, or intoxication. For example, one said, "My head was pounding and I was shaking—I just lashed out before I knew what happened." This aligns with the concept of "unconscious violence" being provoked by acute physical factors (*ResearchBank.ac.nz*).

3. Emotional Overwhelm and Fear: Intense emotions like fear, panic, or desperation emerged in narratives. A participant recalled, "I was terrified, thinking someone was attacking me, so I fought back without thinking." These accounts mirror research linking anxiety and fear to aggression under stress (*ResearchBank.ac.nz*).

4. Reflections on Intent: After the fact, individuals often rationalized their acts as accidents. Many explicitly denied premeditation: "I'm not a violent person; this just came out of nowhere." This theme highlights how perpetrators perceive the violence as disconnected from their usual conscious self. These qualitative themes reinforce that unconscious violence involves an altered state of awareness and intent. They provide face validity to the UVRI approach: the factors we coded (pain, fear, confusion) were frequently mentioned by participants themselves. Quotes illustrating each theme are summarized in Appendix D.

Discussion This study developed a novel threat assessment, the UVRI, and demonstrated its potential to predict "unconscious" violent behavior more accurately than existing tools. The UVRI model achieved high accuracy ($AUC\approx0.86$) and substantial gains in sensitivity/specificity over the HCR-20 and PCL-R benchmarks. These results suggest that integrating behavioral indicators and machine learning can capture risk signals that traditional methods miss (Parmigiani et al.; Menger et al.). The finding that UVRI substantially outperforms HCR-20 supports the notion from prior reviews that ML methods often enhance violence risk prediction (Parmigiani et al.; Menger et al.). Importantly, our feature analysis highlights the variables driving UVRI's success: dynamic, situational risk factors like intense emotion and recent stress. The prominence of these features is consistent with forensic theories; emotion dysregulation and acute stress are recognized pathways to aggressive behavior (Herts et al. 1111; *ResearchBank.ac.nz*).

Comparison with Existing Literature: Our UVRI findings align with research that emphasizes dynamic factors. For instance, Herts et al. showed that emotion regulation difficulties mediate stress-induced aggression in adolescents (1111). We observed a similar mechanism in adults: stressors predicted arousal, which predicted aggression. Likewise, the prominent role of social isolation in UVRI echoes Lankford and Silva's identification of isolation as a common factor in mass shooters. By quantifying such risk domains, the UVRI goes beyond static checklists. Comparatively, prior ML studies reported high performance but often in narrow contexts. Our work differs by explicitly targeting unconscious violence. The UVRI training data included flagged indicators of impairment at the time of assault, which is rarely available. This specialized focus likely contributed to the model's strong accuracy; it is effectively learning the signature of impulsive aggression.

Theoretical Implications: The UVRI findings suggest unconscious violence has discernible precursors. Contrary to the notion that such violence is random or purely inexpressible, our model identifies systematic patterns (e.g. panic, pain) that precede incidents. This supports a view that "unconscious" acts are not wholly unpredictable; they emerge from identifiable risk constellations. Theoretically, this underscores the continuum between conscious and unconscious aggression. Even when intent is absent, certain cognitive-affective processes (heightened threat perception, loss of inhibitory control) are at play. UVRI's success implies these processes can be operationalized and measured.

Practical Implications: For forensic and clinical practice, UVRI could augment existing protocols. For example, mental health staff could use a behavioral checklist (from the UVRI codebook) at patient intake or during care to generate a risk score. A high UVRI score would prompt closer monitoring or de-escalation interventions. This is akin to using HCR-20 for long-term planning, but UVRI is designed for imminent risk. In correctional or emergency settings, automated alerts (from EHR data input) might flag patients at risk of an uncontrolled outburst days in advance, allowing preventive measures. Importantly, UVRI is not meant to replace clinical judgment but to inform it. As with all ML tools, the focus should be on supporting staff decisions by providing objective cues, not dictating them.

Limitations: There are several caveats. The study's sample, though sizable, is not fully generalizable. It was drawn from forensic hospitals and may not reflect community violence. The model's performance might differ in other cultures or systems. Also, while we included many relevant features, unmeasured factors (e.g. hormone levels, fine-grained cognition) could further improve prediction. The qualitative data, while illustrative, came from a subset (n=20) and may not capture all experiences of unconscious violence. Finally, ML models carry risks: overfitting is mitigated by cross-validation, but true external validation is needed.

Ethical Considerations: Predictive tools in violence carry ethical weight. A risk of using UVRI is the potential stigmatization of individuals with certain profiles (e.g. if being socially isolated leads to flagged risk). There is also concern about false positives: incorrectly labeling someone as high-risk could affect their treatment or liberty. Thus, any deployment of UVRI must be accompanied by strict oversight and an emphasis on fairness. For example, transparency in how the model works is crucial (feature importance provides some explainability). Regular audits to check for demographic biases would be prudent. Moreover, UVRI predictions should be used only as one component of a larger assessment, alongside clinician evaluation and legal safeguards. **Future Research:** This proof-of-concept invites further study. Longitudinal research is needed to test UVRI prospectively: does a high score indeed predict future incidents across settings? Also, refining UVRI's features – perhaps through wearable sensors (heart rate) or advanced NLP (to capture tone of speech) – could enhance accuracy. The model could also be adapted to related domains, like domestic violence or school aggression. Crucially, research should engage interdisciplinary collaboration (neuroscience, ethics, AI) to continually validate the conceptual assumptions behind UVRI. In sum, our results demonstrate that by targeting the often-overlooked unconscious dimension of violence and leveraging ML, threat assessment can be meaningfully advanced. UVRI's ability to detect risk beyond conventional metrics may help prevent episodes that current tools would miss, potentially making institutions safer. With further development and ethical integration, such data-driven approaches could transform how forensic and clinical systems intervene in the most dangerous cases.

Conclusion

This study introduced the Unconscious Violence Risk Index (UVRI) and tested its effectiveness in predicting violent incidents lacking conscious intent. By combining coded behavioral indicators with machine learning, the UVRI captured dynamic precursors (emotional dysregulation, acute triggers, isolation) that traditional tools largely omit. The UVRI model demonstrated strong predictive performance (AUC≈0.86), significantly outperforming the HCR-20 and PCL-R on the same cases. Feature analysis and qualitative interviews converged on consistent themes: sudden loss of control under distress and unintended harm. These findings suggest that unconscious or impulsive violence is systematically related to identifiable factors. The UVRI's integration of behavioral science and AI thus provides a novel mechanism for early threat detection. The implications for practice are noteworthy. Mental health and forensic professionals could supplement existing risk assessments with UVRI-based scoring to flag individuals at risk of spontaneous aggression. Early intervention strategies (de-escalation, trauma-informed care) could then be deployed proactively. On the theoretical front, this work underscores that "unconscious" violence need not be incalculable; rather, it often leaves measurable traces in behavior and context. However, caution is advised. Predictive models must be used responsibly. The UVRI should augment, not replace, clinical expertise, and rigorous safeguards should prevent misuse of risk labels. Moreover, our study is an initial exploration; replication and refinement are essential before widespread adoption. Future research should aim to replicate UVRI's success in diverse populations, explore neurobiological correlates of its indicators, and ensure the model's fairness across groups. In conclusion, Predicting Unconscious Violence bridges a critical gap in threat assessment. By scientifically targeting the subtle precursors of unintentional aggression, the UVRI represents a promising advance in public safety and forensic science. It exemplifies how cross-disciplinary innovation—melding behavioral insights with machine learning—can yield powerful tools for violence prevention. Continued development along these lines holds the promise of safer communities and better-informed interventions for those on the brink of violence.

References:

Beech, B., and P. Leather. "Workplace Violence in the Health Care Sector: A Review of Staff Training and Integration of Aggressive Acts in Risk Assessment." Journal of Advanced Nursing, vol. 33, no. 6, 2006, pp. 749–56.

Bohnert, A. M., et al. "Emotion Regulation and Adolescent Aggression." Psychiatric Quarterly, vol. 74, no. 4, 2003, pp. 355–70, https://doi.org/10.1023/A:1026129201939

Braun, V., and V. Clarke. "Using Thematic Analysis in Psychology." Qualitative Research in Psychology, vol. 3, no. 2, 2006, pp. 77–101, https://doi.org/10.1191/1478088706qp063oa

"Deep Learning Models Can Predict Violence and Threats Against Healthcare Providers Using Clinical Notes." *npj Mental Health Research*, vol. 3, no. 61, 2024, https://www.nature.com/articles/ s44184-024-00105-7. Accessed 30 Oct. 2025.

Dobbins, N. J., et al. "Deep Learning Models Can Predict Violence and Threats Against Healthcare Providers Using Clinical Notes." *NPJ Mental Health Research*, vol. 3, no. 61, 2024, https://doi.org/ 10.1038/s44184-024-00105-7.

Douglas, K. S., et al. *HCR-20 V3: Assessing Risk for Violence*. 3rd ed., Mental Health, Law, and Policy Institute, Simon Fraser University, 2013.

"Emotion Dysregulation as a Mechanism Linking Stress Exposure to Adolescent Aggressive Behavior." *PMC*, https://pmc.ncbi.nlm.nih.gov/articles/PMC3448707/. Accessed 30 Oct. 2025.

Fein, R. A., et al. *Threat Assessment: An Approach to Prevent Targeted Violence*. U.S. Department of Justice (NIJ Research in Action), 1995.

"General Violence Risk: A Structured Professional Judgment Approach." *Palo Alto University*, https://paloaltou.edu/resources/business-of-practice-blog/general-violence-risk-hcr-20-v3. Accessed 30 Oct. 2025.

References:

Hare, R. D. The Psychopathy Checklist–Revised. 2nd ed., Multi-Health Systems, 2003.

Herts, K. L., et al. "Emotion Dysregulation as a Mechanism Linking Stress Exposure to Adolescent Aggressive Behavior." Journal of Abnormal Child Psychology, vol. 40, no. 7, 2012, pp. 1111–22, https://doi.org/10.1007/s10802-012-9629-4

"The Impact of Machine Learning in Predicting Risk of Violence: A Systematic Review." *Frontiers in Psychiatry*, vol. 13, 2022, article 1015914, https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2022.1015914/full. Accessed 30 Oct. 2025

Iozzino, L., et al. "Prevalence and Risk Factors of Violence by Psychiatric Acute Inpatients: A Systematic Review and Meta-Analysis." *PLOS ONE*, vol. 10, no. 6, 2015, e0128536, https:// doi.org/10.1371/journal.pone.0128536

Lankford, A., and J. R. Silva. "The Psychological and Behavioral Effects of Social Isolation on Mass Shooters." *Frontiers in Psychiatry*, 2025, https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2025.1686510/full. Accessed 30 Oct. 2025

Ling, X., et al. "A Review of Progress in Violence Risk Assessment Methods." Forensic Science International: Synergy, vol. 1, 2019, pp. 55–72, https://doi.org/10.1016/j.fsisyn.2019.05.001

"Machine Learning Approach to Inpatient Violence Risk Assessment Using Routinely Collected Clinical Notes in Electronic Health Records." PubMed, https://pubmed.ncbi.nlm.nih.gov/31268542/. Accessed 30 Oct. 2025.

McLaughlin, K. A., et al. "Developmental Patterns of Depressive Symptoms." *Journal of Abnormal Psychology*, vol. 120, no. 4, 2011, pp. 825–34.

Menger, V., et al. "Machine Learning Approach to Inpatient Violence Risk Assessment Using Routinely Collected Clinical Notes in Electronic Health Records." *JAMA Network Open*, vol. 2, no. 7, 2019, e196709, https://doi.org/10.1001/jamanetworkopen.2019.6709.

References:

Parmigiani, G., et al. "The Impact of Machine Learning in Predicting Risk of Violence: A Systematic Review." Frontiers in Psychiatry, vol. 13, 2022, article 1015914, https://doi.org/10.3389/fpsyt.2022.1015914

"Psychopathy & Aggression: When Paralimbic Dysfunction Leads to..." *PMC*, https:// pmc.ncbi.nlm.nih.gov/articles/PMC4331058/. Accessed 30 Oct. 2025.

Quinsey, H. J., et al. *Violence Risk Appraisal Guide (VRAG): Predicting Violent Recidivism Among Mentally Disordered Offenders*. Simon Fraser University, 1998.

ResearchBank.ac.nz, ht tps : / /www. researchbank.ac.nz/ server /api /core/bi t s t reams / 07f4bab0-84c1-4c92-8dd7-210b4befe2eb/content. Accessed 30 Oct. 2025.

"A Review of Progress in Violence Risk Assessment Methods." PMC, https://pmc.ncbi.nlm.nih.gov/articles/PMC12396626/. Accessed 30 Oct. 2025.

Shields, A., and D. Cicchetti. "Reactive Aggression in Maltreated Children: The Contributions of Attention and Emotion Dysregulation." *Journal of Clinical Child Psychology*, vol. 27, no. 4, 1998, pp. 381–95.

Springer, U., et al. "Predictors of Violence in Psychiatric Patients." *Psychiatry Research*, vol. 209, no. 3, 2013, pp. 289–98.

"Structured Professional Judgment Tools: A Reference Guide for Counter-Insider Threat Hubs." *Center for Development of Security Excellence*, https://www.cdse.edu/Portals/124/Documents/ jobaids/insider/SPJ-Reference-Guide.pdf. Accessed 30 Oct. 2025.

Swanson, J. W., et al. "A National Study of Violent Behavior in Persons with Schizophrenia." *Archives of General Psychiatry*, vol. 63, no. 5, 2006, pp. 490–99.

"Thematic Analysis." Educational Development, UAMS, https://educationaldevelopment.uams.edu/wp-content/uploads/sites/57/2025/01/9-Thematic_analysis.pdf. Accessed 30 Oct. 2025.

References:

"Threat Assessment: An Approach To Prevent Targeted Violence." National Institute of Justice, https://nij.ojp.gov/library/publications/threat-assessment-approach-prevent-targeted-violence. Accessed 30 Oct. 2025.

Wirrant, I. "Forensic Mental Health Assessment." *Journal of Correctional Psychology*, vol. 2, no. 1, 2016, pp. 21–35.

Zeman, J., et al. "Anger and Sadness Regulation: Predictions to Internalizing and Externalizing Symptoms in Children." *Journal of Clinical Child and Adolescent Psychology*, vol. 31, no. 3, 2002, pp. 393–98.