



Application of Artificial Intelligence in Identifying the Dangerous State of Offenders Using the HCR-20 Instrument: CBT-Based Treatment Strategies

Marzieh Aghasi ¹, Karim Salehi ^{2*}, Norouz Karegari ³

¹ PhD student, Department of Law, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran.

^{2*} Associate Professor, Department of Law, Shahrekord Branch, Islamic Azad University, Shahrekord, Iran
(Corresponding Author)

³ Assistant Professor, Department of Law, Science and Research Branch, Islamic Azad University, Tehran, Iran.

Abstract

Introduction and Objective: The present study aims to develop an integrated conceptual model for applying artificial intelligence (AI) to the interpretation of HCR-20 indicators in order to identify and explain dangerous states among offenders and to design cognitive behavioral therapy (CBT)-based treatment strategies within Iran's criminal justice system.

Methodology: In terms of purpose, this study is basic-applied, and in terms of nature, it is descriptive-analytical. It was conducted through a documentary approach and a systematic review of scientific sources. Data were collected by extracting information from relevant scientific articles, specialized books, theses, and domestic and international legal sources concerning four principal domains: HCR-20, artificial intelligence, CBT, and dangerous states. The data were analyzed qualitatively through qualitative content analysis, comparative analysis, and inferential analysis. Finally, an integrated conceptual model was developed on the basis of the findings of the systematic review.

Findings: The systematic review showed that each of the four principal domains has specific capacities and limitations and that none, independently, can comprehensively address the challenges of identifying dangerous states and designing effective rehabilitative interventions within the criminal justice system. HCR-20, as one of the established violence-risk assessment instruments, has acceptable validity and reliability but faces limitations such as low positive predictive value, limited sensitivity to clinical change, and a lack of randomized controlled trial evidence. AI can increase risk-prediction accuracy by processing large volumes of data and identifying hidden patterns, but it also raises serious concerns regarding algorithmic bias, opacity, accountability, and privacy. CBT, as one of the effective therapeutic approaches, has a positive effect on reducing recidivism, although its use involves challenges such as attenuation of effects over time, compulsory participation, and the gap between assessment and treatment. In Iran's criminal justice system, the absence of a clear legal definition, the lack of a standardized risk-assessment system, weaknesses in correctional and rehabilitation programs, and an emphasis on punitive approaches were identified as major obstacles.

On this basis, the integrated AI-HCR-CBT conceptual model was developed around four principal components—input data, processing and analysis, analytical output, and therapeutic interventions—together with a feedback loop explaining the relationship among AI, HCR-20 indicators, and CBT. With characteristics including personalization, data-driven decision-making, dynamism, transparency, ethical orientation, integration, and adaptability to local conditions, the model can help address existing theoretical and practical gaps and enhance the accuracy, legitimacy, and efficiency of criminal-justice decision-making.

Conclusion: By proposing a comprehensive integrated model, this study takes a step toward evidence-based criminal policy and smarter criminal justice in Iran. The AI-HCR-CBT model can establish a coherent chain consisting of risk assessment through HCR-20, data-driven analysis through AI, and personalized therapeutic interventions through CBT, thereby potentially contributing to reduced recidivism, enhanced public safety, and fairer criminal justice. Implementation, however, requires localization, empirical validation, and appropriate legal, technical, and organizational infrastructure in Iran's criminal justice system. Future studies should empirically validate the model in Iran, examine its application to specific populations, and develop interpretable AI tools for risk assessment.

Keywords: Artificial Intelligence; HCR-20; Dangerous State; Cognitive Behavioral Therapy; Risk Assessment; Recidivism; Criminal Justice

Received: 05 Dec 2023

Received: 29 Dec 2023

Accepted: 31 Dec 2023

1. Introduction

Contemporary criminal justice systems are experiencing unprecedented transformations. The rapid growth of criminal-justice data, the increasing complexity of offenders' behavioral patterns, and the limitations of traditional risk-assessment methods have made it necessary to reconsider the theoretical and practical foundations of managing high-risk offenders. Within this context, the assessment of violence risk and the probability of recidivism has become one of the fundamental issues in modern criminal policy and criminal-justice management. Decisions concerning parole, suspension of punishment, and the level of judicial supervision all depend on an accurate assessment of an individual's potential risk to society. The importance of such assessment lies not only in reducing potential harm to society but also in optimizing correctional resources and improving judicial justice. Moreover, accurate and structured risk assessments contribute to the transparency and legitimacy of judicial decisions by ensuring that decisions are based on objective and reliable data (Douglas et al., 2013). This issue is particularly important in systems confronted with large volumes of criminal cases, because inaccurate assessment may simultaneously undermine justice for defendants and threaten public safety.

In recent decades, instruments based on Structured Professional Judgment (SPJ), particularly HCR-20, have been used as one of the established approaches to violence-risk prediction in various criminal-justice systems (Douglas & Reeves, 2010). By integrating historical, clinical, and risk-management indicators, the instrument seeks to move risk assessment from subjective judgments toward structured, evidence-informed assessment and to provide a reliable basis for judicial and psychological decision-making. Despite these advantages, recent research indicates that even established risk-assessment instruments such as HCR-20 have limitations in accurately predicting violent behavior or recidivism (Douglas et al., 2021). The complexity of the causes of offending, the interaction of individual, psychological, and social variables, and the rarity of extremely violent behavior are among the principal explanations for these limitations (Lipsey et al., 2007).

Accordingly, researchers have increasingly focused on the capacity of artificial intelligence and machine learning to analyze complex criminal-justice data (Parmigiani et al., 2022). These technologies can process large volumes of historical, clinical, and management-related information and identify hidden patterns associated with high-risk behavior, potentially increasing the accuracy of predictions concerning violence and recidivism. Nevertheless, the use of AI to assess and identify dangerous states among offenders is not merely a technical issue; it is also a normative and justice-oriented challenge requiring precise legal, ethical, and conceptual clarification (Yeung, 2018; Završnik, 2021). Furthermore, risk-assessment instruments such as HCR-20 often focus primarily on identifying risk and do not sufficiently address effective therapeutic or correctional interventions (Douglas et al., 2013; Fazel & Singhal, 2022). By contrast, findings from criminal psychology indicate that CBT can substantially reduce the likelihood of recidivism and improve high-risk behaviors (Lipsey et al., 2007; McGuire, 2018). Yet there remains no sufficiently coherent theoretical framework linking risk-assessment results to the design of correctional interventions within criminal justice (Polaschek, 2019; Andrews & Bonta, 2017). This gap reduces the practical value of risk-

assessment systems and limits the effective use of technologies such as AI for planning correctional interventions and managing high-risk behavior.

2. Review of Theoretical Foundations

2.1. Artificial Intelligence

Theoretical definition: Artificial intelligence is a branch of computer science concerned with designing and developing systems capable of simulating human cognitive processes, including learning, reasoning, problem-solving, and decision-making (Russell & Norvig, 2021). In psychology and criminal criminology, AI is used to analyze behavioral data, identify hidden patterns in offenders' behavior, predict the risk of violence or recidivism, and support judicial and correctional decision-making (Zhang et al., 2022; Olawale et al., 2025).

Operational definition: In this study, artificial intelligence refers conceptually to a set of machine-learning algorithms used to analyze HCR-20 data and indicators in order to identify behavioral patterns associated with violence risk, predict the probability of recidivism, and provide the analytical information required to design CBT-based correctional interventions.

2.2. HCR-20 Risk Assessment Instrument

Theoretical definition: HCR-20 is a semi-structured instrument based on guided clinical judgment and designed to assess the risk of violence or recidivism among individuals with criminal histories or psychotic and personality disorders. The instrument contains 20 indicators in three principal domains: historical factors (H), clinical factors (C), and risk-management factors (R) (Douglas et al., 2013). By combining empirical data, clinical observations, and the assessor's professional judgment, HCR-20 provides a structured framework for violence-risk analysis and judicial and therapeutic decision-making.

Operational definition: In this study, HCR-20 is conceptually treated as an independent variable. Its 20 indicators—including a history of violence, emotional instability, and response to treatment—are considered input data for AI-based analyses to identify risk patterns and estimate recidivism. The HCR-20 is also understood in its V3 form as reflecting three temporal dimensions: past, present, and future.

Table 1. Structure and Components of the HCR-20V3 Instrument

Domain	No.	Risk Factor	Nature
Historical (H)	H1	Previous Violence	Static
Historical (H)	H2	Young Age at First Violent Incident	Static
Historical (H)	H3	Relationship Instability	Static
Historical (H)	H4	Employment Problems	Static
Historical (H)	H5	Substance Use Problems	Static
Historical (H)	H6	Major Mental Disorder	Static
Historical (H)	H7	Personality Disorder	Static
Historical (H)	H8	Traumatic Childhood Experiences	Static
Historical (H)	H9	Parental/Family Issues	Static
Historical (H)	H10	Previous Antisocial Behavior	Static

Clinical (C)	C1	Insight	Dynamic
Clinical (C)	C2	Negative Attitudes/Hostility	Dynamic
Clinical (C)	C3	Active Symptoms of Major Mental Disorder	Dynamic
Clinical (C)	C4	Impulsivity	Dynamic
Clinical (C)	C5	Unresponsive to Treatment	Dynamic
Risk Management (R)	R1	Treatment/Supervision Plans	Dynamic
Risk Management (R)	R2	Exposure to Destabilizers	Dynamic
Risk Management (R)	R3	Lack of Personal Support	Dynamic
Risk Management (R)	R4	Stress and Coping	Dynamic
Risk Management (R)	R5	Environmental Instability	Dynamic

The findings indicate that the historical-domain factors (H), as static risk factors, constitute the foundation of risk assessment and provide information about stable behavioral patterns. These factors are empirically and theoretically associated with future violence and are more precisely defined in the third version of the instrument. By contrast, the clinical (C) and risk-management (R) factors are dynamic risk factors that permit changes in risk to be assessed over time and responses to treatment to be monitored.

2.3. Cognitive Behavioral Therapy (CBT)

Theoretical definition: Cognitive behavioral therapy is one of the established therapeutic approaches in clinical psychology and criminal justice and emphasizes the interaction among thoughts, emotions, and behavior. Its central premise is that changing dysfunctional patterns of thinking can correct maladaptive behavior and reduce the occurrence of high-risk or violent behavior (Beck, 2011). In criminal justice, CBT is primarily used to modify distorted beliefs, control anger, increase personal responsibility, and reduce recidivism (Hollin, 2017). In criminal psychology, it is also regarded as an effective intervention for improving self-awareness, impulse control, and reducing antisocial thoughts and tendencies (Lipsey et al., 2019).

Operational definition: In this study, CBT is treated as a dependent variable and conceptually refers to a set of cognitive and behavioral therapeutic interventions whose content is designed and implemented on the basis of HCR-20 risk-assessment results and AI-based data analysis.

2.4. Dangerous States of Offenders

Theoretical definition: A dangerous state refers to a set of psychological and behavioral conditions in which the likelihood of violent, antisocial, or harmful behavior toward others increases (Hanson & Morton, 2009). Such states generally result from complex interactions among personality, emotional, cognitive, and situational characteristics and may be identified through indicators such as impulsivity, emotional dysregulation, paranoid thinking, or lack of empathy.

Operational definition: In this study, dangerous states of offenders are treated as a secondary dependent variable. Operationally, the construct is defined through conceptual analysis of the individual, clinical, and

situational indicators of HCR-20 and interpretation of the results produced by AI models. The level of dangerousness is therefore assessed through theoretical explanation, conceptual inference, and comparison among the historical, clinical, and risk-management components of HCR-20 and their algorithmic interpretation.

2.5. Previous Research

Rahnamā and Mahmoudi (2026), in a study entitled “Intelligentization of Criminal Investigations: Examining Applications of Artificial Intelligence in Crime Detection and Evidence Analysis,” examined the role of AI in crime detection, crime-scene analysis, suspect identification, evidence assessment, and crime prediction. The results indicated that AI can increase the speed and accuracy of data analysis, reduce human error, and enhance predictive capacity, while also creating challenges such as privacy violations, algorithmic bias, and uncertainty concerning accountability.

Nematollahi (2026), in a study entitled “Challenges of Judicial Rules Concerning the Admission and Validity of Digital Evidence and Artificial Intelligence in the Criminal-Proof Process,” examined legal and technical challenges arising from the incorporation of digital evidence and AI systems into criminal proceedings. The findings indicated significant gaps in Iran’s legal system regarding technical standardization of electronic evidence and AI oversight, emphasizing meaningful human oversight and algorithmic interpretability as fundamental conditions for admissibility.

Abdollahi (2026), in a study entitled “The Role of Artificial Intelligence in Predicting the Criminal Profile of Offenders,” analyzed the role of AI in criminal-profile prediction. The findings reported a direct and significant effect coefficient of 0.61 and suggested that police adoption of AI developments could contribute to improving crime-control conditions.

Farber (2026), in a study entitled “Human Experts and Artificial Intelligence Models in Offender Risk Assessment: A Comparative Pilot Study Using HCR-20V3,” compared risk assessments produced by three human experts and three AI models. The results indicated higher risk scores for the AI models (mean 3.4 versus 2.8) and higher inter-rater reliability (ICC = 0.79 versus 0.56). AI systems emphasized historical and static factors, whereas human experts focused more on recent behavioral improvements and rehabilitation potential.

Grenier et al. (2026), in a review of 21 studies concerning the use of natural language processing (NLP) to predict violence in clinical populations, reported that enhanced NLP models consistently outperformed approaches based on structured data, with AUC values above 0.80 for self-selected outcomes and particularly visible improvements over short prediction horizons. However, calibration, interpretability, fairness, and external validation remain barriers to immediate clinical application.

Jiao et al. (2026), in a study of pathways for applying virtual-reality technology to crime-risk prediction, reported that traditional instruments such as questionnaires may have limited ecological validity for individual prediction, whereas virtual reality can simulate immersive crime-related scenarios, capture multimodal data, and improve scalability and cost-effectiveness through effective modeling.

3. Method

The present study is basic-applied in terms of purpose and descriptive-analytical in terms of nature and method. It employs a documentary approach and a systematic review of scientific sources. The data are secondary and were collected by extracting information from scientific articles, specialized books, theses, and relevant domestic and international legal sources. The data were analyzed qualitatively using descriptive-analytical methods to develop a conceptual framework integrating the three principal components of HCR-20, artificial intelligence, and cognitive behavioral therapy (CBT).

4. Findings Related to HCR-20

4.1. Validity and Reliability of HCR-20

To comprehensively evaluate HCR-20 performance, findings from the systematic review and meta-analyses are summarized below.

Psychometric Indicator	Key Findings	Main Sources
Inter-rater reliability	Good to excellent (ICC > 0.75) across different populations	Sea & Bang, 2021; Chen et al., 2023; Hopton et al., 2018
Internal consistency	Optimal internal consistency	Dialnet, 2024
Concurrent validity	Strong correlation with VRS ($r = 0.53\text{--}0.71$), PCL-R, LSI-R, and KORAS-G	Chen et al., 2023; Sea & Bang, 2021
Predictive validity (AUC)	0.69 (0.65–0.72) for Version 2	Ogonah et al., 2023
Predictive validity (Harrell's C)	> 0.77 across different populations	Streb et al., 2025; Chen et al., 2023
Sensitivity to change	Scores decline over time, but the reduction is very small and not associated with reduced recidivism risk	Campbell et al., 2026; O'Shea & Dickens, 2015
Application to women	Clinical subscale is the strongest predictor; further research is needed	Gunenc et al., 2025; Rossdale et al., 2020

The findings show that HCR-20 has good to excellent inter-rater reliability. Chen et al. (2023), in China, studied 152 offenders with mental disorders and found consistently excellent inter-rater reliability for the presence and relevance of risk factors and summary risk ratings (SRRs). Sea and Bang (2021), in a study of 80 adult male offenders in South Korea, also reported good to excellent inter-rater reliability for presence, relevance, and summary risk ratings. A meta-analysis of 25 studies involving 2,517 participants further supported optimal internal consistency (Dialnet, 2024).

Regarding concurrent validity, HCR-20 has demonstrated strong correlations with other established risk-assessment instruments. Chen et al. (2023) reported a strong correlation with the Violence Risk Scale (VRS), ranging from $r = 0.53$ to 0.71 . Sea and Bang (2021) also found good correlations between HCR-20V3 scores and other risk-assessment instruments, including R-LSI, R-PCL, and G-KORAS.

4.2. Limitations of HCR-20

Limitation	Explanation	Implications
Lack of randomized controlled trial evidence	No Level-1 RCT evidence exists for the effectiveness of SPJ in reducing violence (Silva, 2020).	Uncertainty about the instrument's real effectiveness in reducing violence.
Low positive predictive value	Because severe violent events are rare, a large number of false positives are generated (Silva, 2020).	Limitations for high-stakes judicial decisions.

Cognitive biases and conflicts of interest	Narrative fallacies and other heuristic biases may arise; financial conflicts of interest may affect publication (Silva, 2020).	Weakened decision-making and reduced trust.
Limited sensitivity to clinical change	Score reductions over time are very small and are not associated with reduced recidivism risk (Campbell et al., 2026; O'Shea & Dickens, 2015).	Limited ability to monitor treatment progress and make decisions about reducing supervision.
Broad definition of violence	The definition may include verbal threats, which can reduce practical applicability (Silva, 2020).	Large numbers of individuals may be labeled "dangerous."
Variation across populations	Performance varies substantially among women, adolescents, and non-Western cultures (Gunenc et al., 2025; Tolan et al., 2019).	Need for population-specific validation studies.
High risk of bias in studies	Most studies (98%) have been assessed as having a high risk of bias (Ogonah et al., 2023).	Uncertainty regarding the validity of the available evidence.

The most important limitation of HCR-20 is the absence of Level-1 randomized controlled trial evidence demonstrating the instrument's effectiveness in reducing violence. Although numerous studies support its predictive validity, there is limited evidence that using the instrument by itself produces a measurable reduction in violence. As Silva (2020) explicitly notes, there is no Level-1 randomized clinical trial evidence demonstrating that SPJ instruments reduce violence.

4.3. Gaps in the HCR-20 Literature

Research Gap	Current Status	Research Need
Longitudinal studies with RCT design	Lack of randomized controlled trials (Silva, 2020).	Conduct RCTs to establish HCR-20 effectiveness in reducing violence.
Assessment of protective factors	Insufficient attention to protective factors alongside risk factors (Ireland et al., 2024).	Develop integrated models of risk and protective factors.
Application to specific populations	Different performance among women, adolescents, and non-Western cultures (Gunenc et al., 2025; Tolan et al., 2019).	Population-specific validation and localization.
Integration with AI	Insufficient research integrating HCR-20 with machine-learning algorithms.	Develop integrated models to improve predictive accuracy.

Sensitivity to change	Limited evidence linking score reductions to reduced recidivism (Campbell et al., 2026).	Further studies examining changes in scores against real-world outcomes.
Application in Iran's criminal justice system	No standardized risk-assessment system (Mojāhed Pāk-Sheresht & Rezāyati Janbeh-Sarāei, 2025).	Localization and validation of HCR-20 in Iran.

The principal research gaps are: (1) the lack of longitudinal RCTs establishing the effectiveness of HCR-20 in reducing violence; (2) insufficient attention to protective factors alongside risk factors; (3) differential performance in specific populations such as women, adolescents, and non-Western cultures; (4) insufficient research integrating HCR-20 with AI algorithms; and (5) the absence of a standardized risk-assessment system in Iran, which makes localization and validation particularly important. These gaps, especially the limited integration of HCR-20 with AI and the absence of a standardized risk-assessment system in Iran, strongly justify the development of an integrated conceptual model.

5. Findings Related to AI in Risk Assessment

5.1. Capacities and Applications of AI

The systematic review indicates that artificial intelligence and machine learning, as a new generation of risk-assessment technologies, have considerable potential to increase the accuracy, speed, and efficiency of criminal-risk assessment processes.

Application Area	Method/Algorithm	Key Findings	Sources
Recidivism prediction	Combination of CNN and SVM	94% accuracy; AUC = 0.99	Kaveh Qahfarrokhi et al. (2025)
Recidivism prediction	Machine-learning models (systematic review)	Mean accuracy = 0.81; AUC = 0.74	Travaini et al., 2022
Recidivism prediction (big data)	Various ML models on 346,685 individuals	AUC = 0.8660 (general) / 0.8375 (violent)	Springer, 2025
AI versus human experts	LLMs (Claude, ChatGPT, Gemini) within HCR-20V3	Higher risk scores (3.4 vs. 2.8); higher reliability (ICC = 0.79 vs. 0.56)	Farber, 2026
Natural language processing (NLP)	Analysis of free-text clinical notes	AUC > 0.80 for self-selected outcomes; improvement for short-term horizons	Grenier et al., 2026
Dynamic risk-assessment models	Logistic regression, decision tree, random forest	30% increase in accuracy (72% to 82%); AUC = 0.88 and 0.75	Roy & Sehgal, 2025
AI-based risk assessment	COMPAS	Significant racial bias; opaque performance	Oborák, 2025; KCI, 2025

Algorithmic-bias reduction	Hybrid approaches: pre-processing, in-processing, post-processing	Substantial reduction in algorithmic discrimination while maintaining performance	Mistry et al., 2026
----------------------------	---	---	---------------------

The findings show that machine-learning models can perform substantially in predicting recidivism. Kaveh Qahfarrokhi et al. (2025), using a combination of a convolutional neural network (CNN) and support vector machine (SVM), reported 94% accuracy and an AUC of 0.99. A systematic review by Travaini et al. (2022) also reported good performance across machine-learning methods, with mean accuracy of 0.81 and an AUC of 0.74.

5.2. Legal and Ethical Challenges of AI

Challenge	Explanation	Implications	Proposed Solutions
Algorithmic bias and discrimination	Algorithms may reproduce and intensify racial, gender, and social biases (Oborák, 2025; Tolan et al., 2019).	Unfair discrimination against particular groups; violation of equality.	Use balanced datasets, bias-mitigation approaches, and continuous monitoring.
Lack of transparency and interpretability (black box)	Algorithmic decision-making may not be understandable to humans (Oborák, 2025).	Violation of defendants' right to know; reduced public trust.	Develop explainable AI (XAI) and maintain complete documentation.
Accountability	It may be difficult to identify the responsible party when an algorithm makes an error (Lee et al., 2025).	Weakening of the principle of individual responsibility.	Establish a legal framework for allocating responsibility and ensure human oversight.
Privacy violation	Access to sensitive data and possible information leakage (NIST, 2024; Rahnamā & Mahmoudi, 2026).	Violation of citizens' fundamental rights.	Apply data-protection principles, encryption, and restrictions on sharing sensitive data.
Conflict with fair-trial principles	AI-based assessment may conflict with the presumption of innocence, legality principle, and judicial knowledge (Dorosti, 2026; Rezaei & Sharifi-Ārā, 2026).	Erosion of criminal-law human-rights guarantees.	Targeted regulation, meaningful human oversight, and a right to challenge decisions.
Absence of a regulatory framework	Iran lacks sufficiently clear legal and regulatory	Irresponsible and opaque use of AI.	Comprehensive regulation and consideration of the

	frameworks (Nematollahi, 2026; van Dijck, 2022).		European AI Act as a comparative model.
--	--	--	---

The most significant challenge is algorithmic bias and discrimination. Tolan et al. (2019), in research on risk assessment in juvenile justice in Catalonia, found that machine-learning models, despite better predictive performance than SAVRY, showed tendencies toward discrimination against foreigners and certain national groups. Foreign individuals who did not recidivate were almost twice as likely as Spanish individuals to be incorrectly classified as high risk. This occurred even when protected characteristics were excluded from model training. COMPAS is also cited as an example of racial bias in recidivism-risk prediction, with Black individuals being disproportionately classified as higher risk (Oborák, 2025).

6. Key Features of the AI-HCR-CBT Model

Feature	Explanation	Implications
Personalization	Design of treatment programs based on each offender's individual characteristics and identified risk factors.	Increased intervention effectiveness; reduced recidivism.
Data-driven decision-making	Decisions based on objective and measurable data analysis.	Greater accuracy and objectivity in risk assessment.
Dynamism	Updating the model as new data become available, including intervention outcomes.	Continuous improvement in predictive accuracy and intervention effectiveness.
Transparency	Use of explainable AI (XAI) tools to clarify decision-making processes.	Greater legitimacy and ability to challenge decisions.
Ethical orientation	Incorporation of ethical principles—justice, transparency, accountability, and privacy—into model design.	Reduction of algorithmic bias; protection of fundamental rights.
Integration	Combination of the complementary approaches of HCR-20, AI, and CBT within a comprehensive model.	Simultaneous coverage of diagnostic, analytical, and therapeutic dimensions.
Adaptability to local conditions	Ability to adapt the model to the legal and cultural structures of different criminal-justice systems.	Potential implementation in Iran's criminal justice system.

The AI-HCR-CBT model has seven key features that distinguish it from existing models. Personalization permits treatment programs to be designed in accordance with each offender's individual characteristics, consistent with the Risk-Need-Responsivity (RNR) model (Andrews & Bonta, 2017). Data-driven assessment can increase accuracy and objectivity and reduce human error and subjective bias (Travaini et

al., 2022). Dynamism allows predictive accuracy and intervention effectiveness to improve continuously through model updates based on new data (Campbell et al., 2026).

7. Applications of the AI-HCR-CBT Model in Criminal Justice

Application Area	Explanation	Benefits
Judicial decision-making	Provide judges with accurate, evidence-based information about offenders' risk levels for decisions concerning parole, suspension of punishment, and supervision levels.	Greater accuracy and objectivity in judicial decisions; reduced human error and subjective bias (van Dijk, 2022).
Correctional planning	Design personalized therapeutic and correctional programs based on accurate identification of risk factors and treatment needs.	More effective correctional interventions; reduced recidivism (Smith et al., 2024).
Prison management	Classify prisoners according to risk level to optimize resource allocation and prevent violence in correctional environments.	Improved resource management; reduced prison violence (Ireland et al., 2024).
Recidivism prevention	Identify high-risk offenders and provide timely preventive interventions to reduce the likelihood of recidivism.	Reduced recidivism; enhanced public safety (Roy & Sehgal, 2025).
Effectiveness evaluation	Monitor the effectiveness of therapeutic and correctional interventions and continuously update programs based on performance data.	Continuous improvement in intervention quality; increased criminal-justice efficiency (Campbell et al., 2026).
Criminal policymaking	Provide scientific and evidence-based data for developing effective and targeted criminal policies.	Evidence-based criminal policy; increased legitimacy and efficiency of the criminal justice system (Mojāhed Pāk-Sheresht & Rezāyati Janbeh-Sarāei, 2025).

The model can be applied in six key areas of criminal justice. In judicial decision-making, it can provide judges with accurate, evidence-based information about offenders' risk levels for decisions concerning parole, suspension of punishment, and supervision. Such use is intended to increase accuracy and objectivity and reduce human error and subjective bias (van Dijk, 2022). In correctional planning, the model can support personalized therapeutic and rehabilitative programs. In prison management, risk classification may support resource allocation and violence prevention. For recidivism prevention, the model can assist in identifying high-risk individuals and delivering timely interventions. It can also support continuous evaluation of intervention effectiveness and evidence-based criminal policymaking.

8. Conclusion

This study aimed to develop an analytical and conceptual model for applying artificial intelligence to the interpretation of HCR-20 indicators in order to identify and explain dangerous states among offenders and

design CBT-based treatment strategies within Iran's criminal justice system. Using a descriptive-analytical and documentary approach, the study systematically reviewed the literature in four principal domains—HCR-20, artificial intelligence, CBT, and dangerous states—and identified existing theoretical and practical gaps. On this basis, the integrated AI-HCR-CBT conceptual model was developed.

The findings showed that each of the four principal domains has distinct capacities and limitations and that none can independently provide a comprehensive response to the challenges of identifying dangerous states and designing effective correctional interventions. HCR-20 has acceptable validity and reliability but is constrained by low positive predictive value, limited sensitivity to clinical changes, and the lack of randomized controlled trial evidence. AI can process large volumes of data and identify hidden patterns, potentially increasing risk-prediction accuracy, but it also presents serious challenges concerning algorithmic bias, opacity, accountability, and privacy. CBT is an established therapeutic approach with positive effects on recidivism reduction, but it faces challenges including attenuation of effects over time, compulsory participation, and the gap between assessment and treatment.

The integrated AI-HCR-CBT model consists of four principal components—input data, processing and analysis, analytical output, and therapeutic interventions—together with a feedback cycle. It explains how AI, HCR-20 indicators, and CBT can be connected within a coherent framework. The model's characteristics of personalization, data-driven analysis, dynamism, transparency, ethical orientation, integration, and local adaptability may help address theoretical and practical gaps and improve the accuracy, legitimacy, and efficiency of criminal-justice decision-making.

In Iran's criminal justice system, where challenges include the absence of a clear legal definition, the lack of a standardized risk-assessment system, weaknesses in correctional and rehabilitation programs, and an emphasis on punitive approaches, the proposed model may provide a basis for moving from exclusively punitive approaches toward scientific risk management and stronger rehabilitative and reintegrative interventions. Its implementation nevertheless requires localization, empirical validation, and suitable legal, technical, and organizational infrastructure.

By presenting a comprehensive integrated model, the study contributes to the development of evidence-based criminal policy and intelligent criminal justice in Iran. The findings may serve as a basis for future research and practical initiatives and may, subject to empirical validation, contribute to reducing recidivism, enhancing public safety, and promoting fairer criminal justice.

References

1. Babāei, M. A. (2011). *Clinical Criminology*. Tehran: Mizan Publishing.
2. Bonta, J., & Andrews, D. A. (2007). *Risk-Need-Responsivity Model for Offender Assessment and Rehabilitation*. Public Safety Canada.
3. Douglas, K. S., Guy, L. S., & Reeves, K. A. (2021). *The HCR-20V3: Assessing Risk for Violence*. Routledge.
4. Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20 V3: Assessing Risk for Violence – User Guide*. Mental Health, Law, and Policy Institute, Simon Fraser University.
5. Douglas, K. S., & Reeves, K. A. (2010). Historical-Clinical-Risk Management-20 (HCR-20) Assessments of Violence Risk: Predictive Validity and Future Directions. *Criminal Justice and Behavior*, 37(7), 723–740. <https://doi.org/10.1177/0093854810363560>
6. Lipsey, M. W., Landenberger, N. A., & Wilson, S. J. (2007). Effects of Cognitive-Behavioral Programs for Criminal Offenders. *Campbell Systematic Reviews*, 15(1), e1022. <https://doi.org/10.1002/cl2.1022>
7. Parmigiani, G., Barkley, B., Casale, S., Mancini, T., & Fracucchi, S. (2022). The Impact of Machine Learning on Violence Risk Prediction: A Systematic Review. *Frontiers in Psychiatry*, 13, 1015914. <https://doi.org/10.3389/fpsy.2022.1015914>
8. Polaschek, D. L. L. (2019). An Appraisal of the Risk–Need–Responsivity (RNR) Model of Offender Rehabilitation and Its Application in Correctional Treatment. *Legal and Criminological Psychology*, 24(1), 1–20.