

Classification Tree Model for Determining Society Unsafety Factors Convicted

Olha Kovalchuk¹, Nataliia Chudyk¹, Tetiana Drakokhrust¹, Valerii Kaniuka², Roman Ivanytskyi³,
and Larysa Khokhlova³

¹ *Theory of Law and Constitutionalism Department, West Ukrainian National University, 11 Lvivska Str., 46009 Ternopil, Ukraine*

² *Department of Security and Law Enforcement, West Ukrainian National University, 11 Lvivska Str., 46009 Ternopil, Ukraine*

³ *Department of Physics and Mathematics, Ternopil Volodymyr Hnatiuk National Pedagogical University, 2 Maxyma Kryvonosa. Str., 46009 Ternopil, Ukraine*

o.kovalchuk@wunu.edu.ua, n.chudyk@wunu.edu.ua, t.drakokhrust@wunu.edu.ua, valeriikaniuka@gmail.com, romik_iv@ukr.net, larysa_khokhlova@tnpu.edu.ua

Keywords: Justice, Decision Trees, Classification Modeling, Society Safety, Judicial Decision-Making, Informational Support.

Abstract: The global rise in crime rates is a result of the ineffectiveness of the traditional punishment system and the need to develop new approaches to applying relevant punitive measures to criminals, taking into account their level of danger to society. This paper presents a classification Decision Trees model to identify significant factors influencing the objective level of danger a convicted person poses to society. Applied research was conducted based on real data (individual characteristics and criminal history records) of 2,052 convicts serving sentences in correctional facilities in Ukraine. It was found that the most significant predictors for determining the level of danger to society were the number of previous incarcerations and the number of suspended sentences. Assessing the level of societal danger posed by convicts is one of the key aspects of upholding the principles of fair justice. The obtained results can serve as informational support for judicial decisions, ensuring a balance between societal protection and successful offender reintegration.

1 INTRODUCTION

The escalating prevalence of criminal activities across numerous nations has sparked grave apprehensions within societal spheres and posed formidable challenges for law enforcement entities and judicial frameworks. This burgeoning trend poses a palpable threat to society's safety, adversely impacting economic progress, eroding citizens' sense of security, and culminating in an erosion of trust in law enforcement agencies and the judicial branch [1]. The current trajectory necessitates a comprehensive analysis of the underlying factors fueling such negative dynamics and the formulation of effective strategies to counteract this phenomenon at both the international and national levels.

Concurrently, the combat against criminal elements demands a multifaceted approach, which transcends merely augmenting the efficacy of law enforcement agencies but also necessitates

addressing the "prison paradox", wherein an escalation in the incarcerated population does not yield a significant reduction in crime rates and incurs additional financial burdens [2].

Society must acknowledge that not all perpetrators are hardened and irredeemable. Frequently, individuals engage in illicit acts due to a confluence of circumstances, recklessness, or the influence of a deleterious environment. In such instances, it is judicious to differentiate between offenders who do not pose a substantial threat to society and entrenched criminal elements. Furnishing prospects for resocialization and rehabilitation for the former category alleviates the strain on the penitentiary system and paves the way for their reintegration into a law-abiding society [3]. The notion of affording a chance for rectification to certain categories of offenders is pertinent and justifiable.

An objective evaluation of the level of danger a suspect (convict/prisoner) poses to society is a pivotal element in upholding the rule of law, ensuring justice, striking a balance of interests, and enhancing the efficacy of the judicial system [4]. Such an assessment is predicated upon a comprehensive analysis of various individual characteristics to determine an individual's proclivity for recidivism, their social adaptability, and the viability of successful resocialization upon release.

2 RELATED WORKS

The problem of the court's imposition of a fair punishment for a crime has recently attracted the attention of many scholars and practitioners in the legal field [5, 6]. Such punishment should be proportionate to the gravity of the crime committed and take into account all the circumstances and personal characteristics of the accused [7]. The limited number of studies on these issues mostly considers the mental and psychological factors affecting the level of societal danger of convicts [8]. Researchers A. Karlsson and A. Hakansson determined that in addition to well-known risk factors such as male gender and young age, other factors increase the likelihood of offenses. These include the use of amphetamines and injectable drugs, previous convictions for violent and property crimes, as well as homelessness and psychiatric problems [9]. The authors A. Pękala-Wojciechowska et al. proved that physical and mental health is a serious obstacle to the social reintegration of those released from prison [10]. Scholars C. Webster and M. Qasim argued that social and economic problems lead men to commit criminal offenses [11]. However, such studies are partial, and the study of factors from criminal records of previous convictions that are associated with increased societal danger of convicts is extremely rare [12, 13] and requires diverse study and a multidisciplinary approach.

3 METHODOLOGY

One of the key factors in making informed judicial decisions regarding the type and duration of punishment, the possibility of applying a suspended sentence, parole, or probation participation is assessing the level of danger of convicts to society. The empirical basis of our study consisted of real data on 2,052 individuals sentenced to serve their

sentences in penal institutions in Ukraine. For applied research, we used the following variables (individual characteristics of inmates and information about their previous convictions):

- Sex: 1 – male, 2 – female;
- Age: 1 – under 18 years, 2 – from 18 to 30 years; 3 – from 30 to 45 years; 4 – age 45 or older;
- AAP – age at the time of the first conviction to the imprisonment: 1 – under 18 years, 2 – from 18 to 30 years; 3 – from 30 to 45 years; 4 – age 45 or older;
- AAS – age at the time of the first conviction (suspended sentence or the imprisonment): 1 – under 18 years, 2 – from 18 to 30 years; 3 – from 30 to 45 years; 4 – age 45 or older;
- MS – marital status: 1 – single, 2 – married;
- ED – education: 0 – incomplete secondary, 1 – secondary, 2 – special secondary, 3 – incomplete higher, 4 – higher;
- PR – place of residence to the actual degree of punishment: 0 – rural area, 1 – urban area;
- TE – type of employment at the time of conviction to the imprisonment: 0 – unemployed, 1 – part-time, 2 – full-time;
- ED – availability of early dismissals: 0 – no, 1 – yes;
- MD – motivation for dismissal: 0 – no, 1 – yes
- and numerical variables:
- RC – number of the imprisonment;
- SC – number of suspended convictions;

RR – level of danger of convicts to society (low, moderate, high) is the dependent variable, and the other – is the independent variable.

To identify the factors that influence the level of danger to society, we utilized classification using Decision Trees [14].

Formal Decision-Making Criteria: the decision-making process is based on a classification tree method using these specific variables. The decision tree recursively partitions data based on these variables to categorize convicts into three threat levels: low, moderate, and high.

3.1. Tree-Based Predictive Modeling

Decision tree learning is a widely adopted technique in statistics and machine learning domains. The Concept Learning System laid the foundation for all decision tree induction algorithms. These algorithms construct decision trees by employing methods from regression and correlation analysis. One prominent example is the CART (Classification and Regression Trees) algorithm, which recursively partitions the data into two child branches. The splitting criterion

for each branch depends on the number of data points it represents. The split is performed based on the feature that exhibits the highest correlation with the target variable for that branch.

A decision tree has a special root node from which all other nodes can be reached. The terminal nodes are called leaves. Each level of the tree represents a decision point, where a condition is evaluated, and the branches emanating from that node correspond to the possible outcomes of that condition. Essentially, at each decision point, the data instances are sorted such that each instance belongs to exactly one branch. This recursive partitioning process continues until a leaf node is reached, resulting in a hierarchical division of the data into smaller subsets. The number of possible ways to represent decisions can be substantial. Therefore, the choice of the splitting method depends on the type of feature being evaluated and the operation used for condition testing.

This study aims to identify the significant factors for the prisoner's distribution into groups (high, moderate, low) according to the degree of the society's danger. Such studies are being conducted in Ukraine for the first time.

The classification tree method is a flexible tool for predicting the class membership of observations. It provides the ability to classify simultaneously based on multiple variables and in various ways, which facilitates analysis and increases the reliability of results.

3.2. Mathematical Formulation of the Problem

Let: $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ be the training dataset of n observations where $x_i = (x_{i1}, x_{i2}, \dots, x_{im}) \in X$ is the vector of m features (individual characteristics of inmates and information about their previous convictions) $y_i \in Y = \{1, 2, 3\}$ is the class label (level of societal danger: low, moderate, high).

The goal is to find a decision tree function $T: X \rightarrow Y$ that minimizes the misclassification error:

$$T^* = \operatorname{argmin}_{T \in \mathcal{T}} \sum_i L(T(x_i), y_i), \quad (1)$$

where \mathcal{T} is the set of all possible decision trees $L(\hat{y}, y)$ is the 0-1 loss function: $L(\hat{y}, y) = 0$ if $\hat{y} = y$, $L(\hat{y}, y) = 1$ if $\hat{y} \neq y$.

At each node t , the optimal split s^* is chosen to maximize the reduction in impurity:

$$s^* = \operatorname{argmax}_{s \in S} \Delta I(s, t), \quad (2)$$

where $\Delta I(s, t) = I(t) - p_{L} \cdot I(t_{L}) - p_{R} \cdot I(t_{R})$, S is the set of all possible splits $I(t)$ is the impurity measure (Gini or entropy) at node t , p_{L} , p_{R} are the proportions of samples going to left/right child nodes t_{L} , t_{R} are the left and right child nodes.

For entropy calculation at node t :

$$I(t) = - \sum_k p(k|t) \log_2 p(k|t), \quad (3)$$

where $k \in Y = \{1, 2, 3\}$ – class labels (levels of societal danger), $p(k|t)$ – the proportion of samples belonging to class k at node t .

Then the information gain for split s at node t is:

$$IG(s, t) = - \sum_i (n_i/n) I(t_i), \quad (4)$$

where n – is the total number of samples at node t , n_i – the number of samples in child node t_i , $I(t_i)$ – is the entropy of child node i .

The process continues recursively until a stopping criterion is met [15].

4 RESULTS AND DISCUSSION

We applied Tree-Based Predictive Modeling [16] to determine the most significant predictors (individual characteristics of inmates and information about their previous convictions) that influence the level (low, moderate, high) of danger posed by convicts to society. The results of the conducted classification are presented in the form of a tree structure, which branches into left and right branches, each containing 11 nodes (Table 1). The interpretation of the results is significantly simplified by using the classification tree graph (Fig. 1).

In the Node column, the node number is indicated, while in the Size of Node column, the number of objects in the corresponding node is shown.

The left branch contains nodes 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22. The right branch consists of nodes 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23. At the first node, the number of convicts with a high level of danger to society is 492, with a moderate level – 552, and 1,008 convicts are classified as having a low level of danger to society. From node 1, two branches (right and left) emerge with corresponding nodes 2 and 3.

The condition for separating convicts into nodes 2 and 3 is as follows: if the value of RC (number of imprisonments) is less than or equal to 2.5, the level of danger to society posed by the convict is low. 1,008 convicts are classified as having a low level of danger to society, 264 – moderate, and 48 – high.

Table 1: The classification tree structure.

Node No.	Tree structure								
	Dependent variable: RR								
	Left branch	Right branch	Size of node	N in class high	N in class moderate	N in class low	Selected category	Split variable	Split constant
1	2	3	2052	492	552	1008	low	RC	2.5
2	4	5	1320	48	264	1008	low	SC	1.5
4	6	7	1105	0	97	1008	low	RC	1.5
6			796	0	0	796	low		
7	8	9	309	0	97	212	low	SC	0.5
8			212	0	0	212	low		
9			97	0	97	0	moderate		
5	10	11	215	48	167	0	moderate	SC	3.5
10	12	13	185	18	167	0	moderate	SC	2.5
12			136	0	136	0	moderate		
13	14	15	49	18	31	0	moderate	RC	1.5
14			31	0	31	0	moderate		
15			18	18	0	0	high		
11			30	30	0	0	high		
3	16	17	732	444	288	0	high	RC	4.5
16	18	19	454	166	288	0	moderate	SC	1.5
18	20	21	330	42	288	0	moderate	SC	0.5
20			219	0	219	0	moderate		
21	22	23	111	42	69	0	moderate	RC	3.5
22			69	0	69	0	moderate		
23			42	42	0	0	high		
19			124	124	0	0	high		
17			278	278	0	0	high		

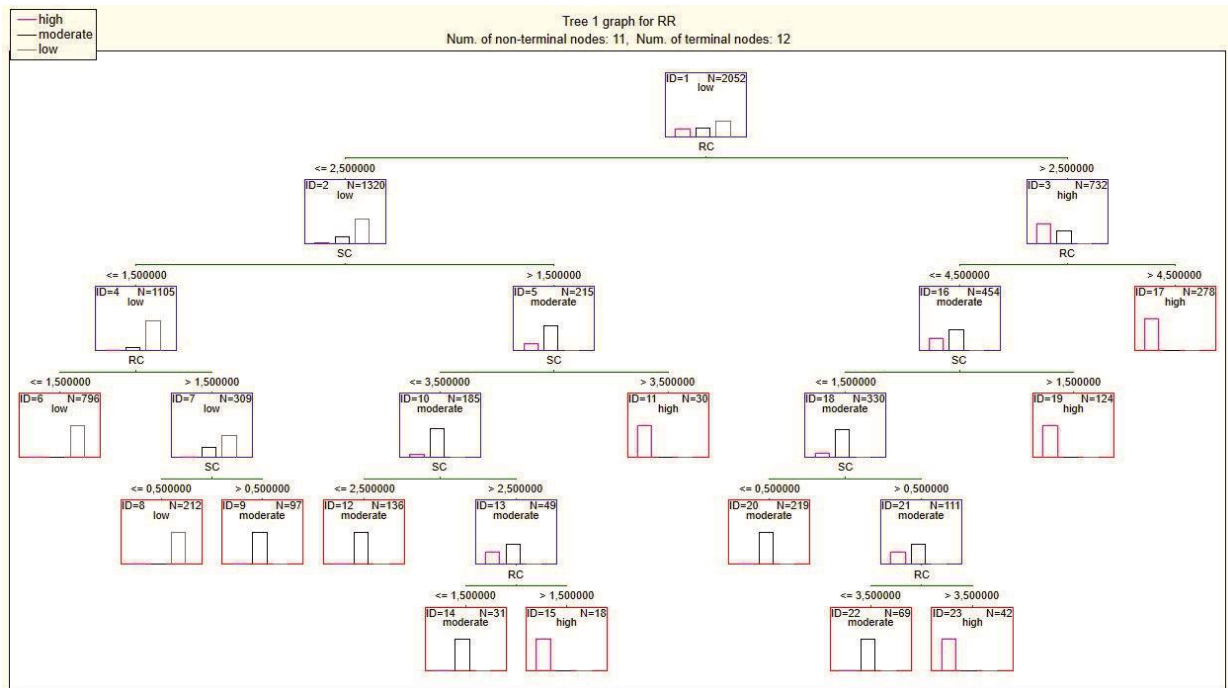


Figure 1: Classification tree.

At nodes 4 and 5, convicts are distributed as follows: if the value of SC (number of suspended convictions) is less than or equal to 1.5, the level of danger to society posed by the convict is low. The classification results at this stage: are 1,008 convicts with a low level of danger to society, 97 – with a moderate level.

For the 1,105 convicts assigned to the group with $SC \leq 1.5$ (node 4), the distribution occurs at nodes 6 and 7 according to the following rule: if $RC \leq 1.5$, then such convicts are finally classified as posing a low level of danger to society (node 6). Otherwise (node 7), the distribution of convicts into groups occurs as follows: if $SC \leq 0.5$ (node 8), the convicts are finally assigned to the group with a low level of danger to society. Otherwise, they are classified as posing a moderate level of danger to society (node 9). Similarly, convicts are distributed into groups based on the level of danger to society at other nodes.

As a result of the conducted classification, out of the 2,052 convicts analyzed, 492 (23.98%) were classified as posing a high level of threat to society, 552 (26.90%) as posing a moderate level of threat, and 1,008 (49.12%) convicts were deemed to have a low level of danger to society. All objects were classified correctly. This is evidenced by the classification matrix (Fig. 2) and the constructed classification matrix diagram (Fig. 3). The classification results showed that the most significant predictors for determining the level of danger to society were the number of previous incarcerations and the number of suspended sentences.

Classification matrix					
Dependent variable: RR					
Options: Categorical response, Analysis sample					
	Observed	Predicted high	Predicted moderate	Predicted low	Row Total
Number	high	492			492
Column Percentage		100.00%	0.00%	0.00%	
Row Percentage		100.00%	0.00%	0.00%	
Total Percentage		23.98%	0.00%	0.00%	23.98%
Number	moderate		552		552
Column Percentage		0.00%	100.00%	0.00%	
Row Percentage		0.00%	100.00%	0.00%	
Total Percentage		0.00%	26.90%	0.00%	26.90%
Number	low			1008	1008
Column Percentage		0.00%	0.00%	100.00%	
Row Percentage		0.00%	0.00%	100.00%	
Total Percentage		0.00%	0.00%	49.12%	49.12%

Figure 2: Classification matrix.

The obtained estimates confirm the results of our previous studies [4, 12] and scientific investigations by other researchers in this field [17, 18]. Moreover, our results correlate with contemporary criminological theories and align with existing knowledge about recidivism risk factors [19, 20].

In this study, we employed the classification tree method to analyze data on 2,052 convicts serving sentences in correctional facilities in Ukraine. The

variables included in the analysis were gender, age, age at first conviction, marital status, education, place of residence, employment, parole availability, and motivation for release, as well as the number of previous incarcerations and suspended sentences. The dependent variable was the level of danger posed by the convict to society (low, moderate, high).

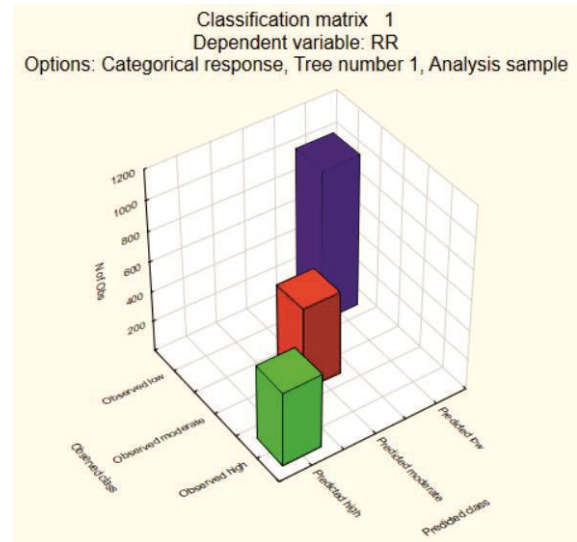


Figure 3: 3-D diagram of the classification matrix.

The integrated threat assessment criterion is primarily based on the number of imprisonments and the number of suspended sentences. The method creates a hierarchical decision tree where:

- $\leq 2-3$ imprisonments suggest low threat;
- $\leq 1-2$ suspended sentences indicate lower societal risk.

Combinations of these factors determine the final threat categorization.

These findings have practical implications for the digitalization of the judicial system and can be used to inform judicial decision-making regarding the type and duration of punishment, the possibility of applying a suspended sentence, parole, or probation participation based on the assessment of recidivism risk to balance societal protection and successful reintegration of offenders.

5 CONCLUSIONS

Assessing the level of danger that convicts pose to society is crucial for making informed judicial decisions regarding the type and duration of

punishment, and the possibility of applying a suspended sentence, parole, or probation participation. The study employed the classification tree method to analyze data from 2,052 convicts serving sentences in correctional facilities in Ukraine. A range of variables were utilized, including gender, age, age at first conviction, marital status, education, place of residence, employment status, availability of parole, motivation for release, as well as the number of previous incarcerations and suspended sentences. The classification results revealed that the most significant predictors for determining the level of danger to society were the number of previous incarcerations and the number of suspended sentences.

These findings have practical implications for the digitalization of the judicial system and can be used to inform judicial decisions regarding the type and duration of punishment, the possibility of applying a suspended sentence, parole, or probation participation based on the assessment of recidivism risk to strike a balance between societal protection and successful reintegration of offenders.

Promising avenues for future research in this area include incorporating additional variables, such as psychological characteristics, history of substance abuse, traumatic past events, and others, to deepen the understanding of factors influencing the risk posed by convicts to society.

REFERENCES

- [1] L.J. Leonard, "Can Restorative Justice Provide a Better Outcome for Participants and Society than the Courts?," *Laws*, vol. 11:14, 2022.
- [2] D. Stemen, "The Prison Paradox: More Incarceration Will Not Make Us Safer," New York: Vera Institute of Justice, 2017, 12 p. Available online: <https://www.vera.org>.
- [3] O.W. Letlape, and M. Dube, "The Parole Officials' Views Concerning the Parole System in Rehabilitating Offenders: Experiences from Brits Community Correction Centre in South Africa," *Soc. Sci.*, vol. 12:410, 2023.
- [4] O. Kovalchuk, M. Kasianchuk, M. Karpinski, and R. Shevchuk, "Decision-Making Supporting Models Concerning the Internal Security of the State," *International Journal of Electronics and Telecommunications*, vol. 69(2), 2023, pp. 301–307.
- [5] O. Arandjelović, "Crime and Punishment: A Rethink," *Philosophies*, vol. 8:47, 2023.
- [6] K. Bullock, and A. Bunce, "The prison don't talk to you about getting out of prison': On why prisons in England and Wales fail to rehabilitate prisoners," *Criminol. Crim. Justice*, vol. 20, 2020, pp. 111–127.
- [7] B. Agozino, "Reparative justice: The final stage of decolonization," *Punishm. Soc.*, vol. 23, 2021, pp. 613–630.
- [8] I. Testoni, I. Nencioni, M. Arbien, E. Iacona, F. Marrella, V. Gorzegno, C. Selmi, F. Vianello, A. Nava, A. Zamperini, et al., "Mental Health in Prison: Integrating the Perspectives of Prison Staff," *Int. J. Environ. Res. Public Health*, vol. 18:11254, 2021.
- [9] A. Karlsson, and A. Håkansson, "Crime-Specific Recidivism in Criminal Justice Clients with Substance Use—A Cohort Study," *Int. J. Environ. Res. Public Health*, vol. 19:7623, 2022.
- [10] A. Pękala-Wojciechowska, A. Kacprzak, K. Pękala, M. Chomczyńska, P. Chomczyński, M. Marczak, R. Kozłowski, D. Timler, A. Lipert, A. Ogonowska et al., "Mental and Physical Health Problems as Conditions of Ex-Prisoner Re-Entry," *Int. J. Environ. Res. Public Health*, vol. 18:7642, 2021.
- [11] C. Webster, and M. Qasim, "The Effects of Poverty and Prison on British Muslim Men Who Offend," *Soc Sci.*, vol. 7:184, 2018.
- [12] K.M. Berezka, O.Ya. Kovalchuk, S.V. Banakh, S.V. Zlyvko, and R. Hrechaniuk, "A Binary Logistic Regression Model for Support Decision Making in Criminal Justice," *Folia Oeconomica Stetinensia*, vol. 22(1), 2022, pp. 1–17.
- [13] C.C. Onyeneke, and A.H. Karam, "An Exploratory Study of Crime: Examining Lived Experiences of Crime through Socioeconomic, Demographic, and Physical Characteristics," *Urban Sci.*, vol. 6, 43, 2022.
- [14] M. Sharma, "Classification using Decision Trees," *Medium*, May 23, 2022. Available online: <https://medium.com/@muditbits/classification-using-decision-trees-6b01939e01a5>.
- [15] C.A. Shannon, "Mathematical Theory of Communication," *The Bell Systems Technical Journal*, vol. 27:1948, pp. 379–423.
- [16] TIBCO Statistica, "User's Guide," 2024. Available online: https://docs.tibco.com/pub/stat/14.0.0/doc/html/Users_Guide/GUID-B3E7379A-0459-4650-A614-1D8974325AB1.html.
- [17] M.R. Kamaluddin, N.A. Mahat, G.A. Mat Saat, A. Othman, I.L. Anthony, S. Kumar, S. Wahab, S. Meyappan, B. Radhakrishnan, and F. Ibrahim, "The Psychology of Murder Concealment," *Acts. Int. J. Environ. Res. Public Health*, vol. 18:3113, 2021.
- [18] S.-C. Hsu, K.-Y. Chen, C.-P. Lin, and W.-H. Su, "Knowledge Development Trajectories of Crime Prevention Domain: An Academic Study Based on Citation and Main Path Analysis," *International Journal of Environmental Research and Public Health*, vol. 19(17):10616, 2022.
- [19] R. Yu, N. Langstrom, M. Forsman, A. Sjolander, S. Fazel, and Ya. Moleru, "Associations between prisons and recidivism: A nationwide longitudinal study," *PLoS ONE*, vol. 17:e0267941, 2022.
- [20] D. Yuhnenko, N. Blackwood, and S. Fazel, "Risk factors for recidivism in individuals receiving community sentences: A systematic review and meta-analysis," *CNS Spectrums*, vol. 25(2), 2020, pp. 252–263.