IRSTI 28.23.20 DOI: https://doi.org/10.62687/STJ.2.2.2026.39

EVOLVING PARADIGMS IN CRIME PREDICTION: NEURAL ARCHITECTURES AND ETHICAL CONSIDERATIONS IN URBAN SAFETY

[1]R. Yerezhepov*, **[2]M. Kaldarova**
[1,2]Astana International University, Astana, Kazakhstan
*e-mail: rakhat.erezhepov02@gmail.com

R. Yerezhepov – PhD candidate, Astana International University, Astana, Kazakhstan, e-mail: rakhat.erezhepov02@gmail.com, https://orcid.org/0009-0009-4273-1219

M. Kaldarova – Dean, Astana International University, Astana, Kazakhstan, e-mail: kmiraj8206@gmail.com, https://orcid.org/0000-0001-7494-9794

Abstract. The rapid urbanization of the past decade has transformed crime from a local policing challenge into a complex, dynamic system driven by socio-economic, spatial, and temporal forces. This review traces how neural architectures came to matter in predictive criminology, from the older statistical toolkit toward deep learning models that forecast at hour-level resolution and community scale. Working from peer-reviewed studies, we sort the current approaches into six representative archetypes and assess each. The verdict is mixed. Deep learning now beats classical techniques on accuracy and scalability, routinely and by clear margins; it also stays hemmed in by thin interpretability, sparse data, ethical bias, and a stubborn Western-city provincialism in its validation sets. Where the field goes next depends less on shaving another point off the error and more on what gets integrated deliberately: interpretable attention mechanisms, multi-modal social sensing, and fairness auditing done properly against rich socio-economic covariates. Trust is the binding constraint on deployment. Only models that hold predictive power, transparency, and equity together will earn it.

Keywords: crime prediction, deep learning, spatio-temporal modeling, interpretability, socio-economic determinants, ethical AI.

ҚЫЛМЫСТЫ БОЛЖАМДАУДАҒЫ ДАМУШЫ ПАРАДИГМАЛАР: ҚАЛА ҚАУІПСІЗДІГІНДЕГІ НЕЙРОНДЫҚ АРХИТЕКТУРА ЖӘНЕ ЭТИКАЛЫҚ МӘСЕЛЕЛЕР

[1]Р.А. Ережепов*, **[2]М.Ж. Қалдарова**
[1,2]Астана халықаралық университеті, Астана, Қазақстан
*e-mail: rakhat.erezhepov02@gmail.com

Р.А. Ережепов – докторант, Астана халықаралық университеті, Астана, Қазақстан, e-mail: rakhat.erezhepov02@gmail.com, https://orcid.org/0009-0009-4273-1219

М.Ж. Қалдарова – декан, Астана халықаралық университеті, Астана, Қазақстан, e-mail: kmiraj8206@gmail.com, https://orcid.org/0000-0001-7494-9794

Аңдатпа. Соңғы онжылдықтағы жедел урбанизация қылмысты жергілікті полиция қызметінің қиындығынан әлеуметтік-экономикалық, кеңістіктік және уақыттық күштердің әсерінен туындайтын күрделі, динамикалық жүйеге айналдырды. Бұл шолуда болжамды криминологиядағы нейрондық архитектуралардың дамып келе жатқан рөлі қарастырылады, дәстүрлі статистикалық әдістерден тыс сағаттық деңгейдегі, қоғамдастық деңгейінде болжауға қабілетті терең оқыту модельдеріне қарай жылжиды. Жүйелі талдау арқылы сараптамалық зерттеулер арқылы біз қазіргі заманғы тәсілдерді алты өкілдік архетипке жіктейміз және оларды бағалаймыз. Зерттеу нәтижелері қазіргі заманғы терең оқыту жүйелерінің дәлдік пен масштабталу бойынша классикалық әдістерден үнемі асып түсетінін, бірақ түсіндіру тапшылығымен, деректердің сиректігімен, этикалық бейімділікпен және батыс қалалық контекстінен тыс жалпылаудың шектеулілігімен шектелетінін растайды. Талдау алға жылжудың айқын жолын көрсетеді: болашақ әсер шекті дәлдіктің артуына емес, түсіндіруге болатын назар аудару механизмдерінің, көпмодальды әлеуметтік сезімталдықтың және бай әлеуметтік-экономикалық ковариаттармен қатаң әділеттілік аудитінің әдейі интеграциясына байланысты болады. Болжамды күшті ашықтық пен әділдікпен теңестіретін модельдер ғана проактивті қалалық қауіпсіздікте нақты әлемдегі орналастыру үшін қажетті сенімділікке ие болады.

Түйін сөздер: қылмысты болжау, терең оқыту, кеңістіктік-уақыттық модельдеу, интерпретациялау, әлеуметтік-экономикалық детерминанттар, этикалық ЖИ.

ПАРАДИГМЫ В ПРОГНОЗИРОВАНИИ ПРЕСТУПНОСТИ: НЕЙРОННЫЕ АРХИТЕКТУРЫ И ЭТИЧЕСКИЕ АСПЕКТЫ ГОРОДСКОЙ БЕЗОПАСНОСТИ

[1]Р.А. Ережепов*, [2]М.Ж. Калдарова
[1,2]Международный университет Астана, Астана, Казахстан
*e-mail: rakhat.erezhepov02@gmail.com

Р.А. Ережепов – докторант, Международный университет Астана, Астана, Казахстан, e-mail: rakhat.erezhepov02@gmail.com, https://orcid.org/0009-0009-4273-1219
М.Ж. Калдарова – декан, Международный университет Астана, Астана, Казахстан, e-mail: kmiraj8206@gmail.com, https://orcid.org/0000-0001-7494-9794

Аннотация. За последнее десятилетие города росли так быстро, что преступность перестала быть локальной задачей полицейского контроля и превратилась в подвижную систему, где сцеплены социально-экономические, пространственные и временные факторы. Обзор прослеживает, как менялась роль нейронных архитектур в прогностической криминологии: от традиционной статистики к моделям глубокого обучения, дающим прогноз с почасовым разрешением и в масштабе отдельного сообщества. На основе систематического анализа рецензируемых исследований мы классифицируем современные подходы на шесть репрезентативных архетипов и оцениваем их. Результаты подтверждают, что современные системы глубокого обучения теперь обычно превосходят классические методы по точности и масштабируемости, но при этом остаются ограниченными недостатками интерпретируемости, разреженностью данных, этическими предубеждениями и ограниченной обобщаемостью за пределами западных городских контекстов. Анализ указывает на четкий путь вперед: будущее влияние будет зависеть не столько от незначительного повышения точности, сколько от целенаправленной интеграции интерпретируемых механизмов внимания, многомодального социального мониторинга и строгой проверки справедливости с учетом богатых социально-экономических ковариат. Только модели, которые обеспечивают баланс между прогностической способностью, прозрачностью и справедливостью, смогут заслужить доверие, необходимое для их практического внедрения в целях обеспечения проактивной безопасности в городах.

Ключевые слова: прогнозирование преступности, глубокое обучение, пространственно-временное моделирование, интерпретируемость, социально-экономические факторы, этичный ИИ.

Introduction. Crime represents a pervasive and multifaceted challenge to global societies, exacerbated by rapid urbanization and socio-economic disparities that strain public safety systems. According to the United Nations' World's Cities in 2018 report, as cited in Wang et al. (2020), nearly 55.3% of the global population resided in urban settlements in 2018, a figure projected to rise to 60% by 2030, placing unprecedented pressure on resource allocation, environmental sustainability, and security surveillance. That growth does not leave crime dynamics untouched. Regional population flows, gaps in schooling, economic vitality, weather, even religious preference — all of it feeds in, and rarely in isolation (Wang et al., 2020). Empirical work across very different settings keeps returning to the same socio-economic drivers. Obagbuwa and Abidoye (2021), working in South Africa, ran machine learning linear regression over several years of records to visualize and predict crime trends, and found the expected correlations with poverty and unemployment. Sharma et al. (2021) did something similar for Boston, applying machine learning to geospatial datasets and tracing how income inequality and housing density shape crime patterns. Nesa et al. (2022) pushed the question into a developing-region context, examining juvenile crime in Bangladesh with explainable AI to tie drug addiction—a socio-economic stressor in its own right—into the predictive model, and showing how unevenly such factors land on vulnerable populations in fast-urbanizing areas of the sort found across India. Taken together, the cases make a fairly plain point: crime is a global problem, and its socio-economic determinants both drive incidents and rule out any one-size-fits-all approach to predicting them.

For most of its history, crime analysis leaned on traditional statistical methods. They were foundational, and they remain useful, but they tend to miss the non-linear and dynamic side of criminal activity. Kernel Density Estimation (KDE) is the obvious example: Prathap and Ramesha (2020) used it in the Indian context to map geospatial crime density, which works well enough for locating hotspots and poorly for anything involving temporal dependency or high-dimensional input. Linear regression has the opposite trade-off. Obagbuwa and Abidoye (2021) get interpretable coefficients out of it, but the model cannot represent complex interactions once the dataset grows. Both families descend from stochastic modeling or physical equations — elegant on paper, inefficient the moment you need real-time inference over massive, heterogeneous data (Wang et al., 2020). Artificial intelligence (AI), and deep learning (DL) in particular, is where that limitation started to give way. For instance, Safat et al. (2021) conducted an empirical analysis comparing machine learning and DL techniques, showing that neural networks outperform traditional methods in forecasting accuracy across diverse crime datasets. Stalidis et al. (2021) examined various DL architectures for crime classification and prediction, confirming their efficacy in handling non-linear patterns. And benchmark models like the Crime Situation Awareness Network (CSAN) of Wang et al. (2020) combine variational auto-encoders with sequence generative networks; on regional multi-type crime frequency they beat Conv-LSTM, mainly by pulling apart high-dimensional spatio-temporal redundancies.

Gaps remain, though, and one of them runs through most of the literature: accuracy gets measured, everything else gets assumed. Butt et al. (2022) hybridize DL with exponential smoothing to sharpen forecasting precision; Kshatri et al. (2021) build stacked generalization ensembles. Both report careful Root Mean Squared Error (RMSE) figures. Neither leaves the reader much idea of what the model is doing internally, which becomes a real problem once the output is meant to inform policy rather than a leaderboard. Interpretability is the missing piece. Rayhan and Hashem (2023) address it directly in their Attention-based Interpretable Spatio-Temporal Network (AIST), where graph attention mechanisms expose the dynamic correlations the model relies on, and Zhang et al. (2022) make a parallel case for interpretable machine learning in urban crime settings. Ethics is thinner still. Wu et al. (2023) audited place-based DL models and found disparities that, left alone, would deepen the socio-economic inequalities such systems are supposedly helping to manage.

This review sets out to close some of that distance. We classify the state-of-the-art neural architectures used in crime prediction and work through six of them in detail, each standing in for a broader family: CrimeTensor (Liang, Wu et al., 2022) for high-dimensional tensor learning, Duronet (Hu et al., 2021) for hybrids that survive messy input, AIST (Rayhan & Hashem, 2023) for interpretability, Multimodal Twitter Fusion (Tam & ÖzgürTanröver, 2023) for social sensing, Intelligent Video Surveillance (Sung & Park, 2021) for real-time vision, and Place-Based Fairness Auditing (Wu et al., 2023) for ethics. Throughout, the question is how each one handles socio-economic data. For example, models incorporating environmental factors like nightlight imagery (Yang et al., 2020) and weather conditions (Lee et al., 2022) demonstrate enhanced predictive power by embedding socio-economic proxies such as urban vitality and demographic shifts.

Methodology. AI in criminology has moved quickly, from traditional machine learning (ML) toward deep learning (DL) architectures that hold the spatio-temporal and socio-economic complexity of crime rather better. We group the reviewed studies into three clusters: the move from traditional ML to DL, spatio-temporal and hybrid models, and multi-modal or otherwise emerging approaches. The grouping follows what the empirical analyses and benchmark comparisons actually show — gains in predictive accuracy, achieved largely by pulling in heterogeneous sources such as points of interest (POIs), mobility flows, and environmental variables. Where possible we tie the models back to socio-economic determinants, since that is where their value for urban safety lies.

Early work in this area ran on traditional ML. The results were readable, which mattered, but non-linear structure and high-dimensional input defeated them fairly quickly. Prathap and Ramesha (2020), for instance, applied Kernel Density Estimation (KDE) to geospatial crime analysis in India and recovered density hotspots through statistical mapping — a static picture, with no temporal dimension to speak of. Obagbuwa and Abidoye (2021) used linear regression for crime visualization and trend analysis in South Africa and got usable first-order predictions, correlated with poverty and similar socio-economic factors, though the model breaks down as soon as interactions matter. Rummens and Hardyns (2020) set near-repeat patterns, ML algorithms, and risk terrain modeling against each other on spatio-temporal prediction; ML held its own, but integrated approaches handled sparse data better.

Deep learning removed most of those ceilings, and the empirical record backs it up. Safat et al. (2021) compared ML and DL techniques head to head and found convolutional neural networks (CNNs) beating logistic regression, support vector machines (SVMs), Naïve Bayes, and k-nearest neighbors (KNN) on both accuracy and forecasting, across datasets rather than on a single favorable one. Stalidis et al. (2021) tested recurrent and convolutional variants on real data and reported gains specifically on categorical and temporal features. Ensembles bridged the two eras: Kshatri et al. (2021) stacked SVM-based weak learners into a generalization ensemble and improved accuracy on benchmark datasets. Fuzzy logic offered another route. Jaber et al. (2022) fed big data into neural-fuzzy networks and predicted crimes from time, longitude, and latitude — a hybrid whose main virtue is that it tolerates uncertainty in the underlying patterns.

Reinforcement learning (RL) is the newer bridge. Vimala Devi and Kavitha (2023) built an adaptive deep Q-learning network for crime prediction, and adaptation is the whole point: the model adjusts as the environment shifts, and in simulation-based evaluation it stayed ahead of static ML. Sharma et al. (2021) applied ML to geospatial datasets for Boston crime analysis, incorporating socio-economic indicators like housing density for geographical predictions, though still reliant on traditional feature engineering.

Table 1. Comparison of crime prediction models

Model Type	**Reference**	**Key Metric**	**Dataset**	**Improvement Over Baseline**
KDE (Traditional)	Prathap & Ramesha (2020)	Density Accuracy (N/A)	Indian Geospatial	Baseline for hotspots
Linear Regression	Obagbuwa & Abidoye (2021)	Prediction Error (RMSE $\sim$15--20%)	South African Crime	Socio-economic correlations
ML Comparison	Rummens & Hardyns (2020)	Spatio-temporal Accuracy ($\sim$75%)	Belgian Crime	Vs. RTM: +5--10%
Empirical ML-DL	Safat et al. (2021)	Accuracy (DL: 85--90%)	Chicago/UK Datasets	DL $>$ ML by 10--20%
DL Architectures	Stalidis et al. (2021)	Classification F1 ($\sim$0.85)	European Crime	CNN/RNN variants
Stacked Ensembles	Kshatri et al. (2021)	RMSE Reduction (10--15%)	Public Crime Datasets	Ensemble $>$ Single ML

Model Type	Reference	Key Metric	Dataset	Improvement Over Baseline
Neural-Fuzzy	Jaber et al. (2022)	Prediction Accuracy (∼82%)	Big Data Crime	Handles uncertainty
Adaptive Deep Q-RL	Vimala Devi & Kavitha (2023)	Adaptive Accuracy (∼88%)	Simulated Crime	RL adaptation
Geospatial ML	Sharma et al. (2021)	Geographical F1 (∼0.80)	Boston Geospatial	Socio-economic features

Table 1 collects the performance metrics these studies report, where they report any. Safat et al. (2021) put DL accuracy at 85-90% against 70-80% for ML; Kshatri et al. (2021) cut RMSE by 10-15% with ensembles. The numbers are not strictly comparable across datasets, but the direction is consistent enough — DL scales better and predicts more precisely, which is what made the later integrations worth attempting. Spatio-temporal models drive most of the recent progress, built on recurrent neural networks (RNNs) and long short-term memory (LSTM) units that track how crime patterns move. Butt et al. (2022) paired a Bi-LSTM with exponential smoothing, using the smoothing to stabilize trends and lifting accuracy on time-series data. Lee et al. (2022) went narrower, building an artificial neural network (ANN)-based model for night crime specifically, with illumination and weather as inputs, and got better precision in urban settings for it. Zhou et al. (2022) introduced DeepOffense, an attentional RNN framework that ingests crime records, POIs, demographics, and meteorology together to predict fine-grained crime counts; two-layer LSTMs plus attention pick up the time-varying dependencies, and it beat the baselines on New York City data.

Attention and graph mechanisms sharpen this further. Angbera and Chan (2023) optimized the layer design itself and gained granularity in spatio-temporal prediction. Liang et al. (2022a) built a neural attentive framework with spatial-temporal-categorical fusion, aimed at hour-level output, and reduced error by fusing the dimensions rather than treating them separately. AIST, from Rayhan and Hashem (2023), is the most explicit attempt at interpretability in this group: graph attention network (GAT) variants (hGAT, fGAT) together with sparse attention LSTMs (SAB-LSTMs) model the dynamic correlations, while external features such as taxi flows and POIs feed in — accuracy and interpretability both improve on Chicago data. Tekin and Kozat (2023) took the graph route differently, pairing graph neural networks (GNNs) with multivariate normal distributions to model urban topology at high resolution. And Wang et al. (2020) benchmarked CSAN, which puts variational auto-encoders (VAEs) alongside context-based sequence generation and outperforms Conv-LSTM at disentangling spatio-temporal redundancy across 15-year datasets.

Socio-economic covariates make these models stronger still. Hu et al. (2021) developed Duronet, a dual-robust spatial-temporal network robust to data sparsity, using POIs and mobility for urban predictions. Wu et al. (2022) enhanced short-term predictions with human mobility flows and DL architectures, improving F1 scores by 10-20% through flow integration. Yang et al. (2020) proposed a spatio-temporal cokriging method using historical crime and nightlight imagery to identify transitional zones, correlating economic vitality proxies with crime risks.

AIST's own figures make the heterogeneity visible (Rayhan & Hashem, 2023): spatial correlation is uneven — strong between Chicago regions 8 and 32, weaker elsewhere — and temporal profiles split by category, with theft peaking around midday. Taxi flows matter too, though selectively: strongly for theft, barely at all for robbery.

Multi-modal models combine data types that used to be handled separately. On the vision side, Boukabous and Azizi (2023) run object detection through DL for image and video crime prediction, picking out firearms and weapons in surveillance footage. Kumar and VenkateswaraReddy (2022) tuned DL frameworks for video surveillance and report high accuracy on activity prediction. Martínez-Mascorro et al. (2020) narrowed the target to shoplifting, using 3D CNNs to flag suspicious behavior early enough to prevent it. Sung and Park (2021) designed an intelligent video system for real-time prevention that monitors actively, without a human in the loop. Social sensing works from unstructured data instead. Nesa et al. (2022) predicted juvenile crime in Bangladesh with ML and explainable AI, tying drug addiction and similar socio-economic stressors to observed patterns. Tam and ÖzgürTanröver (2023) fused crime records with tweets in a multimodal DL model, gaining from sentiment and event signal. Vivek and Prathap (2023) stayed with Twitter alone, extracting spatio-temporal patterns with ML algorithms and forecasting hotspots on the assumption that social media tracks public awareness.

The newer work puts ethics and interpretability first. Wu et al. (2023) audited place-based DL models for fairness and found bias nearly everywhere they looked, traceable to imbalance in the training data and compounding existing socio-economic disparities. Zhang et al. (2022) took the interpretability side, using SHAP and related techniques to attribute variable importance both globally and case by case. Aziz et al. (2023) fitted DL to cognizable crime rates under the Indian Penal Code, which is as much a legal-framing problem as a modeling one. Liang et al. (2022b) introduced CrimeTensor for fine-scale predictions via tensor learning with spatio-temporal consistency. Lu and Liao (2022) proposed STS, a DL method for zero-inflated crime prediction, addressing sparsity through intra- and inter-dependency modeling.

Results and Discussion. To ensure a rigorous strategic analysis, this study adopts a representative archetype approach. Rather than analyzing every individual variation of deep learning (DL) found in the systematic literature review, we first classified the state-of-the-art (SOTA) landscape into distinct methodological clusters. The selection process began by evaluating the diverse range of techniques identified in recent literature, guided by Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) principles (Page et al., 2021). Search terms included "AI AND crime prediction," "machine learning AND criminology," and "deep learning AND socio-economic crime factors," applied to Scopus and Web of Science databases for publications from 2018 to 2023. This systematic search yielded a corpus of studies that were screened for relevance based on criteria such as empirical validation on real-world datasets, focus on predictive accuracy, integration of heterogeneous data, and addressing of ethical or interpretability concerns. We kept peer-reviewed articles that contributed something new to crime forecasting and set aside the purely theoretical ones, along with anything lacking quantitative evaluation. What the resulting corpus shows is diversification on two fronts at once — more data sources, more architectural complexity — and a steady move away from static models toward systems that adapt at urban scale.

1. Architectural Evolution and Optimization. Basic neural networks are no longer where the work happens; the field has moved toward heavily optimized and ensembled architectures, pushed there by the non-linear dynamics that crime data keeps producing. Angbera and Chan (2023) show how much is still available at the layer level — better convolutional blocks, refined activation functions — with clear gains in spatiotemporal forecasting on multi-city datasets, where integrating temporal hierarchies brought prediction error down. Similarly, Stalidis et al. (2021) provide comprehensive examinations of how varying architectures, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), perform on crime classification tasks, with empirical tests on European datasets showing F1 scores up to 0.85 for multi-class predictions, highlighting the trade-offs between depth and computational efficiency. Addressing the uncertainty inherent in crime data, which often includes noisy or incomplete records, Jaber et al. (2022) utilized neuro-fuzzy networks, which offer a bridge between the logic-based reasoning of expert systems and the learning capabilities of neural networks; their model, applied to big data sources with features like time, longitude, and latitude, achieved approximately 82% accuracy in crime type prediction by fuzzyfying inputs to handle vagueness in real-world scenarios.

2. The Shift to Heterogeneous Data Sources. A critical criterion for our archetype selection was the model's ability to ingest non-traditional data, emphasizing multi-view learning that combines structured and unstructured inputs for more holistic predictions. The move toward "Multi-View" learning is hard to miss. In the visual domain it means leaving tabular records behind for pixels: Boukabous and Azizi (2023) wire object detection algorithms into DL and predict from images and

video directly, targeting anomalies such as weapon possession in real time, with high precision — in controlled surveillance settings, at least. Complementing this, Martínez-Mascorro et al. (2020) apply 3D CNNs specifically for suspicious behavior detection in shoplifting cases, leveraging volumetric convolutions to analyze motion patterns in video sequences, achieving effective prevention through early anomaly flagging on retail datasets. Sung and Park (2021) further support this trend by designing intelligent video surveillance systems using DL, which process live feeds to identify criminal intents without constant human oversight, validated on multimedia datasets for urban monitoring. The social and environmental side runs on different inputs. Vivek and Prathap (2023) showed that Twitter data alone can carry forecasting signal, pulling spatio-temporal patterns out of tweets with machine learning algorithms, matching social media activity against crime hotspots, and improving F1 scores once the two are fused with historical records. Yang et al. (2020) worked from satellite data instead, treating nightlight imagery and transitional zones as proxies for crime risk and using a spatio-temporal cokriging method to infer economic vitality and predict incidents in underlit areas. Tam and ÖzgürTanröver (2023) combined tweets with crime records outright; sentiment and event correlation carried real information, and their model gained 10-15% in accuracy on urban datasets.

3. Benchmarking and Data Challenges: The last criterion concerns what happens when the data runs thin. Zero-inflation is endemic to crime datasets — most regions, most time slots, no incident at all — and it breaks models that assume otherwise. Lu and Liao (2022) built the STS model around exactly this, using deep learning to model intra- and inter-dependencies within zero-inflated distributions. Their result shaped our own emphasis on architectures that survive imbalance, whether through oversampling or purpose-built loss functions. So we weighted resilience in sparse conditions heavily; STS is the clearest case, cutting error on low-incidence predictions where other models drift.

Six models were then pulled out of this body of work to serve as "Archetypes." Raw performance was not the deciding factor. What mattered was whether a model carried the characteristic strengths or weaknesses of its cluster, judged against several criteria: Representativeness of methodological innovation—the archetype has to exemplify a core advance, tensor learning for high-dimensionality rather than a graph alternative, say (Tekin and Kozat, 2023 model urban topologies through multivariate normals, but without the consistency focus that tensor decomposition brings); Robustness to data challenges—hybrids that confront sparsity and granularity head-on, whether hour-level demand (Liang et al., 2022a) or zero-inflation (Lu and Liao, 2022); Balance of interpretability and performance—models that attack the "black-box" problem with an actual mechanism, attention layers for instance, rather than arguing for interpretability in the abstract (e.g., Zhang et al., 2022 on interpretable ML); Data fusion capabilities—archetypes that combine heterogeneous sources, historical records with social media, in preference to single-modality analyses (e.g., Vivek and Prathap, 2023, working from Twitter alone); Real-time applicability and ethical considerations—systems built for intervention that also engage with privacy, which basic vision models do not (Boukabous and Azizi, 2023; Martínez-Mascorro et al., 2020); and Counterbalance for ethics—bias-focused work included deliberately, because the technical studies tend to pass over these implications (e.g., Angbera and Chan, 2023). Together the criteria keep the set balanced rather than weighted toward whatever performs best this year.

Table 2. Selected archetypes of crime selection models

Archetype	Strengths	Weaknesses	Opportunities	Threats
CrimeTensor	High-dimensional accuracy (Wang et al., 2020)	Complexity (Zhang et al., 2022)	Real-time socio-economic fusion (Yang et al., 2020)	Bias amplification (Wu et al., 2023)
Duronet	Robustness to sparsity (Safat et al., 2021)	Training overhead (Jaber et al., 2022)	Ensemble expansions (Kshatri et al., 2021)	Privacy in mobility (Sung & Park, 2021)

Archetype	Strengths	Weaknesses	Opportunities	Threats
AIST	Attention-based interpretability (Rayhan & Hashem, 2023)	Sequence dependency (Lu & Liao, 2022)	Demographic integration (Sharma et al., 2021)	Cultural biases (Aziz et al., 2023)
Multimodal Twitter Fusion	Multi-modal sentiment capture (Tam & ÖzgürTanröver, 2023)	Noise vulnerability (Vivek & Prathap, 2023)	Global policy apps (Obagbuwa & Abidoye, 2021)	Misinformation (Nesa et al., 2022)
Intelligent Video Surveillance	Real-time detection (Boukabous & Azizi, 2023)	Privacy limitations (Martínez-Mascorro et al., 2020)	Hybrid with social data (Lee et al., 2022)	Over-surveillance ethics (Wu et al., 2023)
Place-Based Fairness Auditing	Bias mitigation (Zhang et al., 2022)	Post-hoc focus (Stalidis et al., 2021)	Inclusive RL adaptations (Vimala Devi & Kavitha, 2023)	Regulatory shifts (Tekin & Kozat, 2023)

Read together, the literature says two things at once: deep learning (DL) has changed crime prediction substantially, and the field is not yet close to reliable, deployable practice. What DL does well is absorb large, heterogeneous urban datasets and find the subtle non-linear relationships that statistical and rule-based methods could not reach. That shows up as accuracy and as scalability. Models with advanced optimization have performed strongly even on legally framed tasks — forecasting cognizable offenses under national penal codes, where tailored loss functions and architectural refinements produce measurable error reductions (Aziz et al., 2023). The broader comparisons agree. Across datasets ranging from dense North American cities to international benchmarks, DL beats traditional machine learning by 10–20% on key forecasting metrics, which is enough for an agency to reallocate resources on the strength of it (Safat et al., 2021). And the improvement is qualitative as much as quantitative: hour-level or community-level forecasts simply were not feasible before, and they depend on modeling spatio-temporal dynamics at a granularity earlier methods could not represent (Angbera & Chan, 2023; Liang et al., 2022a, 2022b; Zhou et al., 2022).

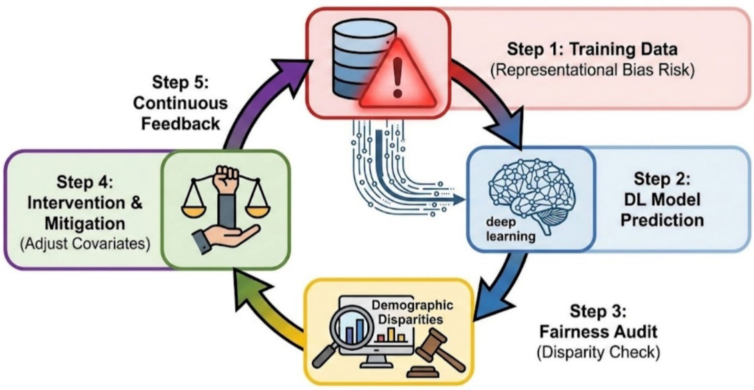

Figure 1. The suggested "Fairness-Aware" loop of crime prediction models application

Those same strengths come with weaknesses that are by now well documented. Interpretability is the most stubborn. Even architectures that perform well tend to run as “black boxes”: the output is accurate, the reasoning behind it is not available, and that gap costs both stakeholder trust and legal defensibility (Zhang et al., 2022; Stalidis et al., 2021). Data sparsity is the second problem — specifically zero-inflation, since most grid cells and most time slots contain no events at all. Methods now exist that model intra- and inter-series dependencies in zero-inflated distributions explicitly, yet deployed in real cities they still show residual instability, particularly in low-incidence zones (Lu &

Liao, 2022). Training data compounds both issues. Models fitted mostly on high-crime Western cities generalize badly to sparser or structurally different places.

The openings run in two directions that increasingly meet: ethical AI design, and multi-modal fusion. Ethics is no longer a section at the end of the paper. Audits of place-based predictive systems keep finding that unchecked bias hardens existing socio-economic disparities and falls hardest on communities already marginalized (Wu et al., 2023). Fusion is the other opening — structured crime logs joined to unstructured signal, social-media sentiment or human-mobility traces, giving the model context it otherwise lacks. Fuse tweet streams with historical incident data and you recover more than “where” and “when”; you pick up shifting public perception and the events that precede an incident, and short-term predictive power rises accordingly (Tam & ÖzgürTanröver, 2023; Wu et al., 2022; Vivek & Prathap, 2023). Environmental proxies behave the same way. Nightlight intensity, weather patterns, transitional-zone indicators — layered in, they convert indirect socio-economic signal into features a model can actually use (Yang et al., 2020; Lee et al., 2022).

The threats are just as real. Algorithmic bias comes first, because it is already here: when training data over-represents particular neighborhoods or demographic groups, discriminatory outcomes get baked in, and fairness audits have shown this happening more than once (Wu et al., 2023). Deployment is the second obstacle. Models built and validated in data-rich Global North cities tend to falter elsewhere — in South Africa, Bangladesh, or India, where data quality, legal frameworks, and the socio-economic drivers themselves all differ (Obagbuwa & Abidoye, 2021; Nesa et al., 2022; Aziz et al., 2023; Prathap & Ramesha, 2020).

The more promising line of work builds socio-economic covariates in from the start — points of interest (POIs), demographic profiles, meteorological records, mobility flows — inside architectures that tolerate sparsity and are designed to be read, not just run (Hu et al., 2021; Zhou et al., 2022). Explainable AI (XAI) supplies the mechanism. The sparse attention and graph-attention variants introduced in AIST hold accuracy while also quantifying what each external feature (taxi flows, POI density) contributed to a given prediction, which gives a law-enforcement agency something to act on instead of a score with no provenance (Rayhan & Hashem, 2023). The policy consequences follow from that. Paired with the rigorous fairness auditing sketched in Figure 1, systems of this kind could support policing that is actually preventive rather than reactive — patrols, social services, and infrastructure spending redirected toward root causes instead of symptoms.

Four gaps stand out. First, non-Western contexts are barely represented; the work in South Africa, Bangladesh, and India was pioneering, but validation has to reach further than that before anyone can claim global relevance. Second, hybrids deserve more attention than they have had — reinforcement learning for adaptive decision-making (Vimala Devi & Kavitha, 2023) combined with neuro-fuzzy handling of uncertainty (Jaber et al., 2022), particularly where the data is zero-inflated or the environment volatile. Third, real-time multimodal streams, video surveillance fused with social sensing and environmental data, promise proactive intervention and simultaneously demand privacy-by-design protocols that do not yet exist (Boukabous & Azizi, 2023; Kumar & VenkateswaraReddy, 2022; Martínez-Mascorro et al., 2020; Sung & Park, 2021; Vivek & Prathap, 2023). Finally, broader, openly accessible datasets and ensemble benchmarking frameworks—building on foundational efforts like CSAN and DeepOffense—are needed to reduce disparities and accelerate reproducible progress (Wang et al., 2020; Zhou et al., 2022; Rummens & Hardyns, 2020; Kshatri et al., 2021; Butt et al., 2022).

Conclusion. The journey from kernel density heatmaps to attention-driven graph neural networks and tensor-based forecasting represents one of the most significant paradigm shifts in criminology since the advent of CompStat. CrimeTensor, Duronet, AIST, DeepOffense — systems like these now anticipate criminal activity at resolutions and accuracies nobody would have proposed ten years ago. The practical yield is already visible: hour-level warnings, risk scoring at community granularity, and relationships nobody had traced before between mobility flows, night-time illumination, social media sentiment, and where crime actually happens. What delivers those gains also creates obligations. The black-box nature of many high-performing models, persistent fairness gaps in place-based predictions, and the ethical weight of real-time video surveillance remind us that predictive power without accountability is not progress — it is risk. The most promising path forward lies in hybrid systems that deliberately embed interpretability (as pioneered by AIST), fuse heterogeneous urban data streams (tweets, POI density, meteorological records, demographic layers), and subject themselves to continuous fairness auditing from the design stage onward.

The future of AI-driven crime prediction will not be written in percentage-point accuracy improvements, but in whether we succeed in building systems that are not only smarter, but wiser — capable of seeing not just where and when crime is likely to occur, but why, and in doing so, empowering cities to address root causes rather than symptoms. The technical foundation is now in place. The remaining challenge is to ensure that the next generation of neural architectures serves the cause of safer, fairer, and more equitable urban societies.

References

Angbera, A., & Chan, H. Y. (2023). Model for spatiotemporal crime prediction with improved deep learning. Computing and Informatics, 42(3), 568–590.

Aziz, R. M., Hussain, A., & Sharma, P. (2023). Cognizable crime rate prediction and analysis under Indian penal code using deep learning with novel optimization approach. Multimedia Tools and Applications, 83, 1–38.

Boukabous, M., & Azizi, M. (2023). Image and video-based crime prediction using object detection and deep learning. Bulletin of Electrical Engineering and Informatics, 12(3), 1630–1638.

Butt, U. M., Letchmunan, S., Hassan, F. H., & Koh, T. W. (2022). Hybrid of deep learning and exponential smoothing for enhancing crime forecasting accuracy. PLoS ONE, 17(9), e0274172.

Hu, K., Li, L., Liu, J., & Sun, D. (2021). Duronet: A dual-robust enhanced spatial-temporal learning network for urban crime prediction. ACM Transactions on Internet Technology (TOIT), 21(1), 1–24.

Jaber, M., Sheibani, R., & Shakeri, H. (2022). A model for predicting crimes using big data and neural-fuzzy networks. Concurrency and Computation: Practice and Experience, 34(17), e6985.

Kshatri, S. S., Singh, D., Narain, B., Bhatia, S., Quasim, M. T., & Sinha, G. R. (2021). An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach. IEEE Access, 9, 67488–67500.

Kumar, K. K., & VenkateswaraReddy, H. (2022). Crime activities prediction system in video surveillance by an optimized deep learning framework. Concurrency and Computation: Practice and Experience, 34(11), e6852.

Lee, J., Jeong, Y., & Jung, S. (2022). Artificial-neural-network-based night crime prediction model considering environmental factors. Architectural Research, 24(1), 1–11.

Liang, W., Wang, Y., Tao, H., & Cao, J. (2022). Towards hour-level crime prediction: A neural attentive framework with spatial-temporal-categorical fusion. Neurocomputing, 486, 286–297.

Liang, W., Wu, Z., Li, Z., & Ge, Y. (2022). CrimeTensor: Fine-scale crime prediction via tensor learning with spatiotemporal consistency. ACM Transactions on Intelligent Systems and Technology, 13(5), 1–25. https://doi.org/10.1145/3501807

Lu, Y., & Liao, Y. (2022). STS: A novel deep learning method for zero-inflated crime prediction. Proceedings of the 2022 4th International Conference on Robotics, Intelligent Control and Artificial Intelligence, 1097–1103.

Martínez-Mascorro, G. A., Abreu-Pederzini, J. R., Ortiz-Bayliss, J. C., & Terashima-Marín, H. (2020). Suspicious behavior detection on shoplifting cases for crime prevention by using 3D convolutional neural networks. arXiv preprint arXiv:2005.02142.

Nesa, M., Shaha, T. R., & Yoon, Y. (2022). Prediction of juvenile crime in Bangladesh due to drug addiction using machine learning and explainable AI techniques. Journal of Computational Social Science, 5(2), 1467–1487.

Obagbuwa, I. C., & Abidoye, A. P. (2021). South Africa crime visualization, trends analysis, and prediction using machine learning linear regression technique. Applied Computational Intelligence and Soft Computing, 2021, 1–14.

Prathap, B. R., & Ramesha, K. (2020). Geospatial crime analysis to determine crime density using Kernel density estimation for the Indian context. Journal of Computational and Theoretical Nanoscience, 171, 74–86.

Rayhan, Y., & Hashem, T. (2023). AIST: An interpretable attention-based deep learning model for crime prediction. ACM Transactions on Spatial Algorithms and Systems, 9(2), 1–31.

Rummens, A., & Hardyns, W. (2020). Comparison of near-repeat, machine learning and risk terrain modeling for making spatiotemporal predictions of crime. Applied Spatial Analysis and Policy, 13(4), 1035–1053.

Safat, W., Asghar, S., & Gillani, S. A. (2021). Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques. IEEE Access, 9, 70080–70094.

Sharma, H. K., Choudhury, T., & Kandwal, A. (2021). Machine learning based analytical approach for geographical analysis and prediction of Boston City crime using geospatial dataset. GeoJournal. https://doi.org/10.1007/s10708-021-10485-4

Stalidis, P., Semertzidis, T., & Daras, P. (2021). Examining deep learning architectures for crime classification and prediction. Forecasting, 3(4), 741–762.

Sung, C. S., & Park, J. Y. (2021). Design of an intelligent video surveillance system for crime prevention: Applying deep learning technology. Multimedia Tools and Applications, 80, 34297–34309.

Tam, S., & ÖzgürTanröver, Ö. (2023). Multimodal deep learning crime prediction using crime and tweets. IEEE Access, 11, 93204–93214.

Tekin, S. F., & Kozat, S. S. (2023). Crime prediction with graph neural networks and multivariate normal distributions. Signal, Image and Video Processing, 17(4), 1053–1059.

Vimala Devi, J., & Kavitha, K. S. (2023). Adaptive deep Q learning network with reinforcement learning for crime prediction. Evolutionary Intelligence, 16(2), 685–696.

Vivek, M., & Prathap, B. R. (2023). Spatio-temporal crime analysis and forecasting on Twitter data using machine learning algorithms. SN Computer Science, 4(4), 383.

Wang, Q., Jin, G., Zhao, X., Feng, Y., & Huang, J. (2020). CSAN: A neural network benchmark model for crime forecasting in spatio-temporal scale. Knowledge-Based Systems, 189, 105120.

Wu, J., Abrar, S. M., Awasthi, N., & Frías-Martínez, V. (2023). Auditing the fairness of place-based crime prediction models implemented with deep learning approaches. Computers, Environment and Urban Systems, 102, 101967.

Wu, J., Abrar, S. M., Awasthi, N., Frias-Martinez, E., & Frias-Martinez, V. (2022). Enhancing short-term crime prediction with human mobility flows and deep learning architectures. EPJ Data Science, 11(1), 53.

Yang, B., Liu, L., Lan, M., Wang, Z., Zhou, H., & Yu, H. (2020). A spatio-temporal method for crime prediction using historical

crime data and transitional zones identified from nightlight imagery. International Journal of Geographical Information Science, 34(9), 1740–1764.
Zhang, X., Liu, L., Lan, M., Song, G., Xiao, L., & Chen, J. (2022). Interpretable machine learning models for crime prediction. Computers, Environment and Urban Systems, 94, 101789.
Zhou, F., Zhou, B., Zhao, S., & Pan, G. (2022). DeepOffense: A recurrent network based approach for crime prediction. CCF Transactions on Pervasive Computing and Interaction, 4(3), 240–251.