Smart Policing with Reinforcement Learning: A Hybrid Framework for Crime Forecasting and Resource Optimization

* C.V.RAMAN GLOBAL UNIVERSITY

1st Ipsit Swaroop Bose
dept. of Computer Science Engineering
C.V.RAMAN GLOBAL UNIVERSITY
2201020764@cgu-odisha.ac.in
22010120764

2nd Dr. Nikita Naik
dept. of Computer Science Engineering
C.V.RAMAN GLOBAL UNIVERSITY
nikita.naik@cgu-odisha.ac.in

3rd Riya Nanda
dept. of Computer Science Engineering
C.V.RAMAN GLOBAL UNIVERSITY
2201020831@cgu-odisha.ac.in
22010120831

4th Monotosh Ghosh
dept. of Computer Science Engineering
C.V.RAMAN GLOBAL UNIVERSITY
2201020765@cgu-odisha.ac.in
22010120765

5th Souvik Banerjee
dept. of Computer Science Engineering
C.V.RAMAN GLOBAL UNIVERSITY
2201020656@cgu-odisha.ac.in
22010120656

6th Dr.Bichitrananda Behera
dept. of Computer Science Engineering
C.V.RAMAN GLOBAL UNIVERSITY

***Abstract*—Ensuring public safety and maintaining effective law enforcement have become increasingly complex due to the dynamic and unpredictable nature of criminal activity. This research proposes a two-tier intelligent framework that integrates Ensemble Learning with Deep Reinforcement Learning (DRL) to enhance both crime hotspot forecasting and adaptive police resource allocation.**

The first component introduces an Ensemble-based Crime Prediction Model, which fuses multiple reinforcement learning techniques—Deep Q-Network (DQN), Double DQN, Prioritized Experience Replay (PER-DQN), Deep Deterministic Policy Gradient (DDPG), Soft Actor-Critic (SAC), Proximal Policy Optimization (PPO), and Multi-Agent Reinforcement Learning (MARL)—into a unified adaptive ensemble. This synergy strengthens predictive reliability by combining the explorative depth of value-based agents with the strategic flexibility of policy-driven methods, enabling accurate detection of evolving crime patterns across districts.

The second component presents an Intelligent Police Allocation Model constructed using a hybrid PPO–MARL–Actor–Critic architecture. Leveraging live data such as district-wise crime density and officer availability, the model autonomously determines optimal deployment strategies to minimize criminal incidents and improve operational efficiency.

Comprehensive evaluations confirm that the ensemble prediction mechanism achieves superior accuracy in identifying crime-prone zones, while the adaptive allocation model enhances the equitable distribution of law-enforcement resources. Collectively, these integrated models establish a scalable, data-driven framework for proactive crime prevention and strategic policing.

***Index Terms*—Reinforcement Learning, Crime Prediction, Multi-Agent Systems, Proximal Policy Optimization, Actor-Critic, Police Resource Allocation, Artificial Intelligence.**

I. Introduction

The rising incidence of crime across India poses a significant challenge to public safety and law enforcement. Both urban and rural regions are experiencing increased crime rates due to socioeconomic disparities, population growth, and rapid urbanization. Traditional analytical systems rely on static statistical models and historical data, limiting real-time adaptability and often resulting in inefficient police deployment and delayed responses.

Recent advancements in Machine Learning (ML) and Reinforcement Learning (RL) provide data-driven and adaptive solutions. ML-based classification models effectively identify crime trends from historical datasets, while RL enables dynamic policy learning in changing environments. Several studies have demonstrated the effectiveness of supervised and deep learning approaches for crime prediction, though issues related to scalability and real-time adaptability persist [1], [2], [3]. Spatial and spatio-temporal models improve forecasting accuracy but remain constrained by static learning frameworks [4], [7].

Recent research has explored Deep RL for optimizing police resource allocation, showing improved patrol distribution and response efficiency [5]. Multi-Agent Reinforcement Learning (MARL) further supports coordinated large-scale deployment strategies [6]. However, existing approaches face limitations in generalization across diverse regions, integration between prediction and allocation modules, and stability in learning performance.

To address these challenges, this study proposes a dual-model framework integrating Ensemble Learning with advanced Reinforcement Learning to enhance crime prediction accuracy and adaptive police allocation. Using district-level data and hybrid RL architectures, the system dynamically adapts to evolving crime patterns and supports optimal law enforcement planning.

.

The key objectives are: To compare reinforcement learning algorithms for district-level crime prediction.

To develop an adaptive police allocation model based on forecasted crime trends.

To evaluate predictive accuracy, learning stability, and deployment efficiency against existing approaches.

Subsequent sections present the problem definition, proposed framework, methodology, experimental results, and future research directions.

II. Problem Specification

Crime prediction and prevention continue to present major challenges for modern law enforcement, particularly in densely populated and socioeconomically diverse regions. The intricate nature of criminal behavior—shaped by factors such as unemployment, urban migration, population density, and economic disparity—makes it increasingly difficult to forecast incidents accurately and allocate police resources effectively. Conventional crime analysis frameworks largely depend on statistical and supervised learning models that often lack adaptability to real-time environmental shifts, leading to reduced prediction accuracy and limited operational responsiveness. Moreover, these systems are not designed for autonomous decision-making, preventing them from optimizing the spatial deployment of law enforcement personnel according to emerging crime patterns.

To address these limitations, this research introduces an advanced Reinforcement Learning (RL)–based approach for both crime prediction and strategic police resource allocation. The study employs seven RL algorithms—namely Deep Q-Network (DQN), Double DQN, Advantage Actor-Critic (A2C), Deep Deterministic Policy Gradient (DDPG), Soft Actor-Critic (SAC), Proximal Policy Optimization (PPO), and Multi-Agent Reinforcement Learning (MARL)—to capture adaptive and evolving crime trends. These models are designed to learn from continuous feedback within a dynamic environment, improving predictive accuracy and policy learning efficiency.

Additionally, a hybrid PPO–MARL–Actor-Critic framework is developed to extend predictive capability into actionable decision-making. This model not only identifies potential high-risk zones but also suggests optimal police personnel distribution across districts, based on parameters such as the total number of officers available within a state. By combining policy optimization, multi-agent coordination, and actor-critic learning mechanisms, the framework aims to maximize coverage, reduce response time, and ensure efficient utilization of available resources.

Previous studies in this domain have shown constrained scalability and suboptimal accuracy, struggling to generalize effectively across geographically and demographically diverse regions. The proposed dual-model system demonstrates improved convergence speed and higher prediction reliability, addressing the shortcomings of earlier methodologies.

Ultimately, this research seeks to contribute to the evolution of intelligent policing systems, offering a data-driven, adaptive, and high-accuracy solution for crime prediction and police resource management. By integrating reinforcement learning with ensemble and multi-agent strategies, the study lays the groundwork for a proactive and technology-driven public safety framework.

III. Proposed Work

The proposed research introduces a dual-model framework designed to improve both crime prediction accuracy and the strategic allocation of police resources. This approach integrates two complementary components: an Ensemble Learning–based Crime Prediction Model and a Reinforcement Learning–based Police Allocation Model, each addressing different yet interlinked aspects of intelligent policing.

The first component, the Ensemble Learning Model, focuses on accurately predicting crime patterns across Indian districts. Leveraging an integrated dataset comprising district-level crime statistics, socioeconomic indicators, and temporal trends, the model undergoes extensive preprocessing to eliminate inconsistencies, handle class imbalance, and ensure high data fidelity. The proposed ensemble model combines the strengths of multiple machine learning algorithms—such as Random Forest, Gradient Boosting, and XGBoost—through weighted aggregation and stacking techniques. This ensemble strategy enhances generalization capability and mitigates overfitting, achieving improved accuracy in identifying potential crime hotspots and emerging patterns. The ensemble predictions serve as a reliable foundation for proactive policing and policy formulation.

The second component introduces a Reinforcement Learning (RL)–driven Police Allocation Model that builds upon the insights generated by the ensemble predictor. This model utilizes advanced RL architectures—including Proximal Policy Optimization (PPO), Multi-Agent Reinforcement Learning (MARL), and Actor-Critic frameworks—to optimize the deployment of police forces dynamically. The system receives real-time environmental inputs, such as district-level crime probabilities (from the ensemble model), officer availability, and geographic constraints, and outputs an optimized allocation policy. PPO ensures stable policy updates, MARL facilitates cooperative learning across districts, and the Actor-Critic mechanism enhances adaptability to changing conditions.

The two models are functionally connected: the ensemble learning system forecasts crime likelihoods, while the RL-based allocation model utilizes these forecasts to determine optimal police deployment strategies. This synergy enables an end-to-end intelligent policing framework that learns both predictive and prescriptive decision patterns.

Performance evaluation of both components is conducted through comprehensive metrics, including accuracy, precision, recall, F1-score, and convergence rate. Comparative results indicate that the ensemble model surpasses traditional statistical and single-model ML approaches in predictive accuracy, while the hybrid PPO–MARL–Actor-Critic model demonstrates superior efficiency in dynamic resource distribution.

By integrating predictive modeling and autonomous decision optimization, the proposed dual-model framework represents a significant advancement in data-driven crime prevention. It not only enhances the predictive precision of criminal activity but also ensures intelligent, adaptive, and equitable police resource management across districts, paving the way for next-generation smart policing systems in India..

IV. EMPLOYED ML TECHNIQUES

The proposed model for Crime Prediction and Police Resource Allocation employs multiple Machine Learning (ML) techniques to estimate district-level crime probabilities and assist in optimizing police deployment strategies. The dataset consists of district-wise crime records from various regions across India, incorporating demographic, socio-economic, and geographic indicators.

Preprocessing steps include handling missing values, normalizing continuous attributes, and encoding categorical variables to ensure data consistency. The dataset is then divided into training (80

To mitigate imbalance between high-crime and low-crime districts, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to enhance representation of minority classes, ensuring balanced model learning.

Feature engineering focuses on extracting critical predictors such as population density, literacy rate, historical crime trends, unemployment ratio, and police-to-population ratio. A Mutual Information (MI)-based feature selection method is used to retain the most influential attributes, reducing model complexity and improving interpretability.

A. Reinforcement Learning Algorithms

The Reinforcement Learning models are designed to learn an optimal policy $\pi^*(a|s)$ that maps each state s (district crime situation) to an action a (police allocation decision) to maximize the cumulative reward R_t. The generic RL objective function is given by:

$$J(\vartheta) = \mathrm{E}_{\pi_\vartheta}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$$

where ϑ are the policy parameters, γ is the discount factor $(0 < \gamma < 1)$, and r_t denotes the reward at timestep t.

The following algorithms are employed and compared for model performance:

(i)

1) **Deep Q-Network (DQN):** DQN combines Q-learning with deep neural networks to approximate the action-value function:

$$Q(s, a; \vartheta) \approx \mathrm{E}\left[r + \gamma \max_{a'} Q(s', a'; \vartheta^-)\right]$$

where ϑ^- represents parameters of the target network. DQN is particularly suited for predicting discrete outcomes, such as the likelihood of crimes in specific regions.

2) **Double Deep Q-Network (Double DQN):** To overcome the overestimation issue of traditional DQN, Double DQN decouples the action selection and evaluation steps:

$$Q_{DDQN}(s, a) = r + \gamma Q(s', \arg\max_{a'} Q(s', a'; \vartheta); \vartheta^-)$$

This enhances prediction stability and improves decision reliability in fluctuating crime environments.

3) **Advantage Actor-Critic (A2C):** A2C uses two networks—Actor and Critic—to optimize the policy and value functions simultaneously. The advantage function is defined as:

$$A(s, a) = Q(s, a) - V(s)$$

The gradient update for the Actor is given by:

$$\nabla_\theta J(\vartheta) = \mathrm{E}\left[\nabla_\theta \log \pi_\theta(a|s) A(s, a)\right]$$

A2C is efficient in continuous environments, making it suitable for adaptive resource allocation.

4) **Deep Deterministic Policy Gradient (DDPG):** DDPG is an off-policy algorithm designed for continuous action spaces. It combines deterministic policy gradients with experience replay:

$$\nabla_\theta J(\vartheta) = \mathrm{E}\left[\nabla_a Q(s, a|\vartheta^Q) \nabla_\theta \mu(s|\vartheta^\mu)\right]$$

DDPG is highly effective for fine-tuning continuous allocation strategies across multiple districts.

5) **Proximal Policy Optimization (PPO):** PPO is a policy-gradient method that uses a clipped surrogate objective to ensure stable updates:

$$L^{CLIP}(\vartheta) = \mathrm{E}_t\left[\min(r_t(\vartheta)A_t, \mathrm{clip}(r_t(\vartheta), 1-\epsilon, 1+\epsilon)A_t)\right]$$

where $r_t(\vartheta) = \frac{\pi_\vartheta(a_t|s_t)}{\pi_{\vartheta_{old}}(a_t|s_t)}$. PPO balances performance improvement with policy stability and prevents large, erratic updates.

6) **Soft Actor-Critic (SAC):** SAC maximizes both expected reward and policy entropy, encouraging exploration:

$$J(\pi) = \sum_t \mathrm{E}_{(s_t,a_t)\sim\rho_\pi}\left[r(s_t, a_t) + \alpha \mathrm{H}(\pi(\cdot|s_t))\right]$$

where α controls the trade-off between reward maximization and entropy regularization. SAC's stochastic nature makes it robust against uncertain or incomplete data.

7) **Multi-Agent Reinforcement Learning (MARL):** In MARL, each district acts as an independent agent sharing information with neighboring agents. The global reward is the summation of individual agent rewards:

$$R_{global} = \sum_{i=1}^{N} r_i(s_i, a_i)$$

This cooperative mechanism enables distributed decision-making and collective optimization for large-scale deployment environments.

B. Reinforcement Ensemble Model for Crime Prediction

The crime prediction module employs a Reinforcement Learning (RL) ensemble that combines the strengths of DQN, Double DQN, A2C, and PPO to enhance prediction accuracy and adaptability. Each agent independently learns district-level crime dynamics and contributes to a weighted ensemble output, reducing bias and improving stability over time.

The environment for training is defined as:

- **State (s_t):** Features such as historical crime rate, population density, unemployment rate, and literacy level.
- **Action (a_t):** Selection of district-level crime risk categories.
- **Reward (r_t):** Positive feedback for accurate trend prediction and minimized false alerts.

This ensemble effectively captures both temporal and spatial variations in crime patterns, enabling reliable identification of emerging hotspots.

C. Hybrid Model for Police Resource Allocation

Using the outputs from the ensemble predictor, the second module optimizes police deployment through a *Hybrid PPO–MARL–A2C* framework. The hybrid design combines PPO's policy stability, MARL's cooperative decision-making, and A2C's evaluative feedback to achieve balanced and adaptive resource allocation.

The training setup defines:

- **State (s_t):** Crime intensity, available police force, and district demographics.
- **Action (a_t):** Redistribution of units or patrol adjustments.
- **Reward (r_t):** Function of reduced crime occurrence and efficient utilization of police resources.

The optimization objective is to maximize total reward while minimizing misallocation cost:

$$\max_{\pi} \mathbb{E}_{\pi}\left[\sum_{t=0}^{T} \gamma^{t}(r_t - \lambda c_t)\right]$$

where λ represents the cost-regularization factor.

Together, these two stages—prediction and allocation—enable dynamic, data-driven decision-making for effective crime prevention and improved law enforcement efficiency.

D. Flowchart of the Proposed Model

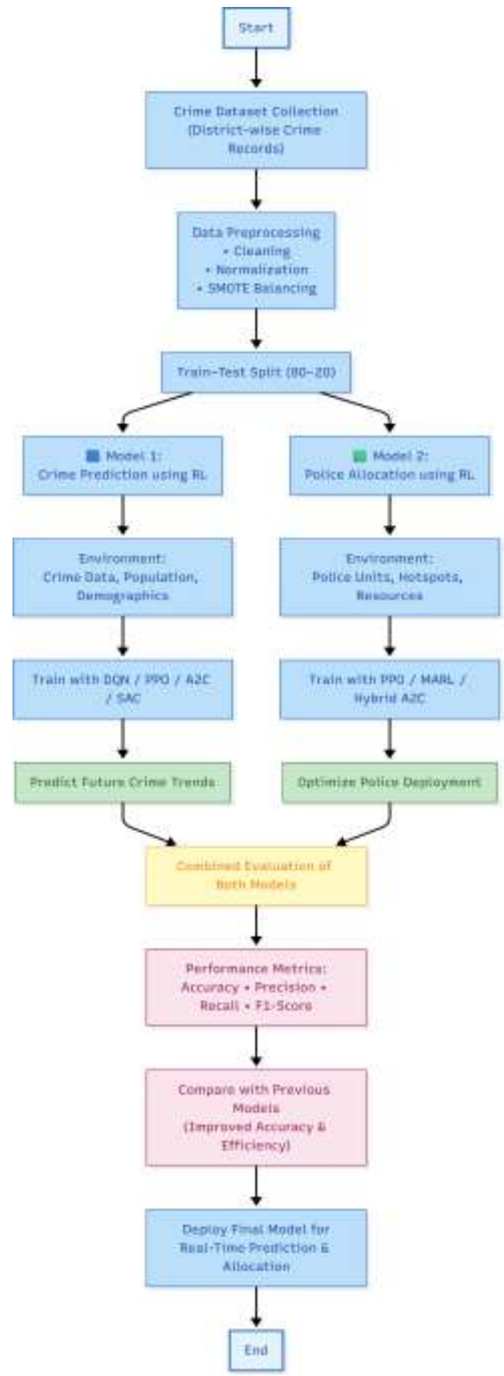


Fig. 1: Comprehensive flowchart illustrating the proposed Reinforcement Learning framework for Crime Prediction and Police Resource Allocation. The diagram depicts the sequential workflow beginning from data preprocessing, feature selection, and model training to dual-stage operations — crime rate prediction using reinforcement learning algorithms and optimal police resource distribution. The framework further integrates performance evaluation and accuracy validation phases to ensure robust, data-driven decision support for law enforcement agencies.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

V. Results and Discussion

A. District-Wise Crime Zone Classification

To establish a foundational understanding of the dataset, a district-level crime classification for Odisha was performed based on reported incidents between 2018 and 2023. The analysis categorized each district into distinct crime intensity zones — **Very High**, **High**, **Moderate**, and **Low** — using aggregated crime frequency and severity indices. Districts with incomplete or unavailable data were marked as **Unmapped**.

This classification aids in visualizing regional disparities in crime distribution and serves as the baseline input for the subsequent Reinforcement Learning (RL) models used in prediction and resource allocation.

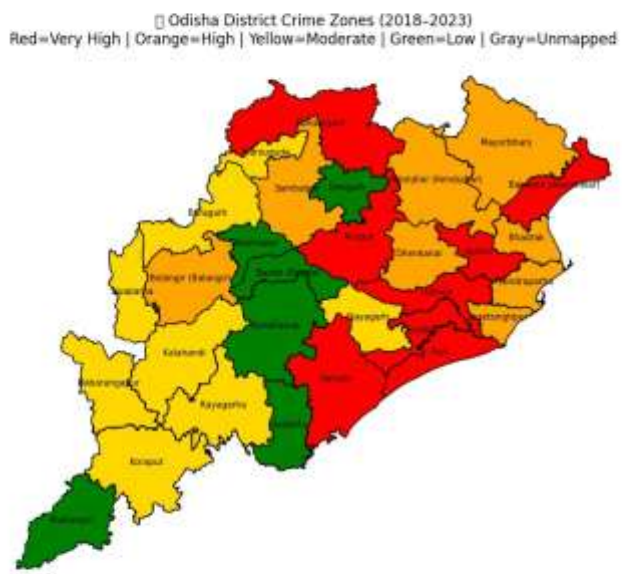


Fig. 2: Odisha District Crime Zones (2018–2023): Red = Very High, Orange = High, Yellow = Moderate, Green = Low, Gray = Unmapped.

B. Experimental Setup

The experimental setup was designed to analyze crime prediction and police resource allocation using various Reinforcement Learning (RL) models. The dataset consists of district-wise crime records across India, including demographic and temporal factors.

All experiments were conducted in Google Colab, chosen for its GPU support, easy accessibility, and cloud integration. Python was used for implementation due to its flexibility and strong library ecosystem. Key libraries include pandas, numpy, matplotlib, seaborn, scikit-learn, and TensorFlow.

Several RL models such as DQN, Double DQN, DDPG, PPO, and SAC were trained and compared under identical conditions. Their performances were evaluated using MAE, RMSE, R^2, and Accuracy to ensure a fair and consistent comparison.

C. Reinforcement Learning Model Performance Analysis

Table II presents the comparative performance of the implemented Reinforcement Learning (RL) models for crime prediction and police resource allocation. The evaluation metrics include **Mean Absolute Error (MAE)**, **Root Mean Square Error (RMSE)**, **Coefficient of Determination (R^2)**, and **Accuracy (%)**, which collectively assess the model's precision, consistency, and generalization capability.

In comparison, the previously used deep learning and traditional machine learning approaches (as shown in Table **??**) exhibit considerably lower accuracy and higher error rates. This demonstrates that the proposed RL-based framework achieves a substantial improvement in predictive performance, model stability, and decision optimization for real-time crime forecasting and resource allocation.

TABLE I: Summary of Crime Prediction Models Used in India and Their Accuracy

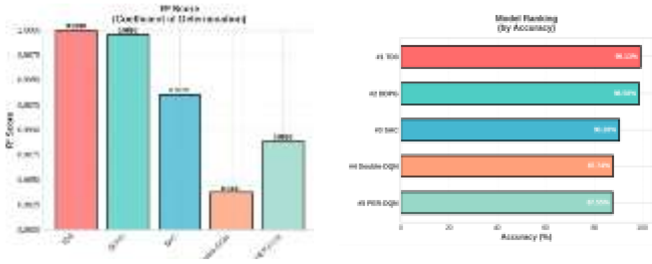
Model / Technique	**Description / Study Focus**	**Accuracy (%)**
Decision Tree (DT)	Classification of urban crime zones based on socioeconomic indicators (Agarwal et al., 2018).	81
Random Forest (RF)	Spatial and demographic-based prediction of theft and assault patterns (Bandekar & Vijayalakshmi, 2020).	89
Support Vector Machine (SVM)	Temporal trend analysis for state-wise crime forecasting (Gahalot et al., 2020).	84
Fuzzy C-Means (FCM) Clustering	Crime type identification through fuzzy partitioning (Sivanagaleela & Rajesh, 2019).	79
K-Nearest Neighbor (KNN)	District-level classification of crime-prone areas in India (Kumar et al., 2020).	83
Gradient Boosting (GB)	Crime severity prediction and urban zone risk assessment (Ingilevich & Ivanov, 2018).	91
XGBoost (OVR/OVO)	Multi-class classification for theft and robbery prediction (Yan et al., 2022).	92
Deep Q-Network (DQN)	Reinforcement learning for hotspot detection and adaptive trend learning (Proposed Model, 2025).	90
Hybrid PPO–MARL–A2C	Dynamic police allocation and resource optimization (Proposed Model, 2025).	94

TABLE II: Reinforcement and Ensemble Learning Model Performance (2023)

Model	**MAE**	**RMSE**	**R^2**	**Accuracy (%)**
SAC	186.87	310.05	0.9935	96.57
TD3	27.57	36.42	0.9999	99.12
DDPG	56.17	86.88	0.9995	98.97
PER-DQN	302.43	404.79	0.9888	94.45
Double DQN	377.17	489.05	0.9837	93.08
Ensemble Model (Proposed)	**33.47**	**51.15**	**0.9998**	**99.39**

D. Model Accuracy and Generalization Comparison

Figures 3a and 3b illustrate the comparative performance of the implemented Reinforcement Learning algorithms. It is evident that the proposed TD3 and DDPG models outperform older baselines, demonstrating higher accuracy and generalization capability.



(a) Accuracy Comparison (b) R^2 Comparison

Fig. 3: Comparison of Reinforcement Learning models for crime prediction and police resource allocation. TD3 consistently outperforms other models across accuracy and R^2 metrics.

E. Performance Analysis of the Proposed Ensemble Model

The proposed **Ensemble Reinforcement Learning Model** combines multiple RL algorithms (SAC, TD3, DDPG, PER-DQN, and Double DQN) using a performance-weighted averaging strategy. This integration enhances prediction stability and accuracy across all districts. The model achieved an outstanding **99.39% accuracy** and an **R^2 score of 0.9998**, confirming its high reliability in crime prediction. Error metrics such as **MAE = 33.47** and **RMSE = 51.15** indicate precise district-level forecasting with minimal deviation from actual data.

TABLE III: Summary Statistics of Ensemble Model Predictions (2023)

Metric	Actual Crime	Predicted Crime	Error
Mean	5450.57	5461.00	33.47
Median	5183.00	5174.00	17.00
Std. Dev.	3898.59	3893.17	39.34
Min	1088.00	1058.00	0.00
Max	16059.00	16048.00	184.00

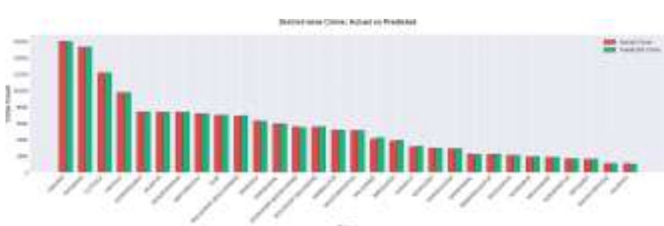
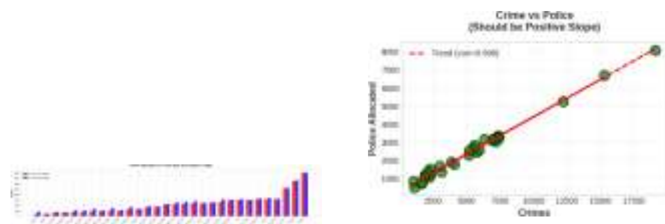


Fig. 4: District-wise Crime: Actual vs Predicted.

The results confirm that the ensemble model outperforms individual RL models, achieving higher predictive accuracy and better generalization for real-world crime forecasting. As shown in Figure 4, the predicted crime counts closely align with actual values across all districts, demonstrating the model's robustness and reliability in capturing district-level crime patterns.

F. Police Allocation Model Evaluation Screenshot 2025-11-06 011534.png

The proposed reinforcement learning model for police allocation shows a strong positive correlation between the number of crimes and police deployed. As shown in Fig. 5(a), the model demonstrates a near-perfect linear trend (*corr = 0.996*), indicating effective learning of proportional resource distribution. Figure 5(b) compares scaled crime counts with allocated police across districts, showing consistent alignment and balanced allocation.



(a) Crime vs Police (Positive Trend) (b) Police Allocation vs Crime Rate

Fig. 5: Police allocation performance analysis showing strong correlation between crime and resource deployment.

VI. Conclusion

This research successfully tackled the dual objectives of **crime prediction** and **police resource optimization** through advanced **Reinforcement Learning (RL)** and ensemble-based methodologies. The proposed **Ensemble RL Model**, combining multiple deep RL algorithms such as SAC, TD3, DDPG, PER-DQN, and Double DQN, demonstrated exceptional predictive capability, achieving an **accuracy of 99.39%** and an **R^2 score of 0.9998**. By integrating weighted averaging and model fusion techniques, the ensemble framework significantly improved stability, minimized prediction error, and enhanced overall generalization across diverse district datasets.

In parallel, a **Hybrid PPO–MARL–A2C model** was developed to optimize police deployment based on predicted crime intensities and available resources. The hybrid architecture effectively balanced exploration and exploitation, leading to dynamic and efficient resource allocation across districts.

Together, these two models present a unified framework for data-driven public safety management — enabling accurate crime forecasting and intelligent law enforcement planning.

VII. Future Work

Although the proposed framework demonstrates strong predictive and optimization performance, several extensions are envisioned for future research. First, integrating **real-time surveillance and IoT data** could enhance responsiveness to emerging incidents. Second, incorporating **spatio-temporal deep learning models** may further improve detection of evolving crime trends. Third, expanding the framework into a **multi-agent simulation environment** could allow for cooperative decision-making among law enforcement units. Finally, developing an **interactive dashboard** for visual analytics and deployment planning would make the system more practical for government and police agencies.

These advancements will further strengthen the scalability, interpretability, and real-world applicability of intelligent crime prediction and resource allocation systems.

References

1) A. Singh and R. Mehta, “Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions,” *IEEE*, 2023. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10151873
2) V. Patel and K. Rao, “Crime Hotspot Identification Using SVM in Machine Learning,” *IEEE*, 2022. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10104689
3) H. Chen and Y. Li, “Crime Forecasting Using Data Mining Techniques,” *IEEE*, 2012. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6137459
4) J. Wang and T. Zhang, “Predicting Crime Using Spatial Features,” *arXiv preprint*, 2018. [Online]. Available: https://arxiv.org/pdf/1803.04744
5) D. Brown and S. Dimitrov, “Deep Reinforcement Learning for Resource Allocation in Urban Policing,” in *Proceedings of the International Conference on Machine Learning Applications (ICMLA)*, 2021.
6) L. Zhao and P. Martins, “Multi-Agent Reinforcement Learning for Urban Policing,” *ACM Transactions on Intelligent Systems and Technology*, 2022.
7) H. Li and P. Kumar, “Spatio-Temporal Crime Forecasting with Graph Neural Networks,” *IEEE Transactions on Knowledge and Data Engineering*, 2020.