HYBRID MACHINE LEARNING–REGRESSION FRAMEWORK FOR VALIDATING A MULTIDIMENSIONAL CRIME INDEX ON CRIMES AGAINST WOMEN

[1]Dr. G. R. Anantha Raman, [2] P.Y.N.S. MANOJ ,[3] Y.SRILATHA,[4] K.VINEETH REDDY,[5] B.SUNIL

[1] Professor, Department of Computer Science & Engineering (Artificial Intelligence & Machine Learning), Malla Reddy College of Engineering , Hyderabad, India.

[2,3,4,5] Students, Department of Computer Science & Engineering (Artificial Intelligence & Machine Learning), Malla Reddy College of Engineering, Hyderabad, India.

ABSTRACT

Crime against women remains a persistent social, legal, and security challenge across regions worldwide [5], [6]. Traditional crime metrics often oversimplify the multifaceted nature of gender-based violence, resulting in weak predictive insights and incomplete policy recommendations [8], [12]. This study proposes a hybrid analytical model combining Machine Learning (ML) classification techniques with regression-based statistical validation to construct and validate a Multidimensional Crime Index (MDCI), an approach aligned with earlier efforts to fuse spatial analytics and ML for crime research [1], [3], [4]. The model integrates diverse socio-economic, demographic, and spatiotemporal variables to more accurately reflect crime severity, vulnerability, and community-level risk, consistent with prior studies that emphasize contextual risk factors [7], [11], [13], [14]. Experimental results demonstrate that the hybrid framework improves prediction accuracy, robustness, and interpretability when compared with standalone ML or regression methods [2], [3]. The proposed index shows potential as a decision-support tool for law enforcement, policymakers, and NGOs focused on women's safety [5], [9].

Keywords: Crime Against Women, Multidimensional Crime Index (MDCI), Machine Learning (ML), Regression Analysis, Hybrid Analytical Model, Gender-Based Violence, Predictive Crime Analytics,

I. INTRODUCTION

Crimes against women—including harassment, domestic violence, trafficking, and sexual assault—have escalated in many regions despite advancements in legal frameworks [5], [6], [9]. These crimes are often influenced by structural inequalities, socio-economic conditions, and cultural norms, making them complex and multidimensional in nature [8], [11], [13]. Existing crime indices often rely on aggregated statistics that fail to sufficiently capture context-specific factors, thereby limiting predictive precision and practical applicability, a limitation noted in earlier spatial and hotspot-based investigations [1], [4], [7]. A multidimensional crime index capable of encoding and quantifying various determinants of crime is necessary for deeper insight into gender-based violence [3], [12]. However, validating such an index requires robust analytical techniques that can manage heterogeneous variables. Traditional regression methods offer interpretability but struggle with nonlinear interactions, while machine learning techniques excel at pattern recognition but may lack transparency—a challenge highlighted in prior hybrid-modeling studies [2], [3], [14]. Combining the two has the potential to enhance both prediction accuracy and analytical depth.

This research presents a hybrid ML–regression framework to validate an MDCI specifically designed for analyzing crimes against women. By combining supervised ML models with regression-based validation, the study aims to deliver an empirically reliable, interpretable, and scalable tool for crime analysis [3], [7]. The expected outcome is a policy-oriented index that supports preventive interventions, resource allocation, and risk mapping, reinforcing the need for evidence-driven approaches to women's safety [5], [9], [13].

II. Literature Survey

1. Author: Sharma et al. (2019) – "Socio-Economic Determinants of Crime Against Women"

Sharma et al. examined the relationship between socio-economic variables—including literacy, employment, and urbanization—and the prevalence of crimes against women in developing regions. Their research demonstrated that these variables strongly influence reporting patterns and actual crime incidence. The study highlighted the need for multidimensional indicators rather than traditional single-metric crime rates. The authors emphasized that classical regression methods effectively identify statistically significant predictors but struggle when nonlinear interactions become dominant. They observed that complex social factors such as cultural norms and family structures require analytical models capable of capturing latent variables and hidden correlations. The study concluded that integrating socio-economic indicators into crime analysis improves predictive accuracy but recommended an expanded modeling framework involving advanced data-driven approaches. This recommendation aligns with the need for hybrid ML-regression models for validating holistic crime indices.

2. Author: Banerjee & Roy (2020) – "Machine Learning Models in Crime Pattern Prediction"

Banerjee and Roy focused on the application of ML techniques—including Random Forest, SVM, and Gradient Boosting—to predict crime hotspots in metropolitan regions. Their findings indicated that ML models outperform traditional statistical approaches in handling large datasets with mixed variables.

However, they noted that many ML models lack interpretability, making them insufficient for policy-level decision-making, where explanations are crucial for societal and government acceptance. In the context of crimes against women, where transparency is essential, relying solely on ML can be problematic.

The authors recommended the development of hybrid systems combining the predictive capabilities of ML with the interpretative strength of regression models. Such systems could bridge the gap between accuracy and explainability—an approach central to this study.

3. Author: Singh & Thomas (2021) – "Multidimensional Indices for Social Vulnerability Analysis"

Singh and Thomas proposed a multidimensional index to assess community vulnerability to various social issues, including crime. Their model incorporated environmental, demographic, and socio-economic dimensions, showing that composite indices provide richer insights compared to single-dimensional statistics.

They emphasized the importance of rigorous validation of multidimensional indices to ensure accuracy and credibility. Their work identified regression-based validation as a key tool for verifying index reliability and assessing the significance of each dimension. The study acknowledged the potential of ML to enhance index validation by discovering non-obvious patterns in vulnerability data. They suggested that future research should explore hybrid techniques—precisely the direction undertaken in the present study.

4. Author: Patel et al. (2022) – "Deep Learning for Gender-Based Violence Analysis"

Patel et al. applied deep learning models to analyze textual and multimedia evidence of gender-based violence. Their work demonstrated that advanced neural networks can detect emotional cues, contexts of abuse, and patterns in reported incidents with high accuracy. Despite promising results, the authors acknowledged limitations including high computational cost, dependency on large training datasets, and reduced interpretability. For regions where crime reporting is inconsistent, reliance on deep learning alone is insufficient. The authors recommended integrating interpretable models—such as linear or logistic regression—to validate deep learning outcomes. Their conclusion further supports the utility of hybrid approaches in crime analytics.

5. Author: Verma & Kulkarni (2023) – "Statistical Modelling of Crime Trends Through Regression"

Verma and Kulkarni investigated long-term crime trends using time-series regression models. They found that regression provides strong explanatory value, helping identify underlying drivers of crime fluctuations such as migration, unemployment, and demographic shifts.

However, they also noted regression's limitations in handling high-dimensional datasets with complex interactions. These challenges restrict regression's ability to capture nonlinear dynamics inherent in crimes against women. The authors concluded that incorporating ML techniques could overcome these limitations by enhancing predictive accuracy while retaining regression-based interpretability. This supports the rationale for developing a hybrid ML-regression validation framework.

III. EXISTING SYSTEM

Existing crime analysis systems largely rely on historical data, descriptive statistics, and regression-based assessments. These systems aggregate crime rates but do not integrate multidimensional factors such as socio-economic indicators, behavioral variables, spatiotemporal patterns, and demographic attributes. Machine learning efforts exist but often operate independently, lacking integration with statistical validation. Moreover, crime indices used by government bodies are generally simplistic, making them unsuitable for predicting risk or validating complex indicators for crimes against women. These limitations lead to incomplete assessments and limited usability for preventive policymaking.

IV. PROPOSED SYSTEM

The proposed system introduces a **Hybrid Machine Learning and Regression Framework** for validating a **Multidimensional Crime Index (MDCI)** tailored to crimes against women. ML models—including Random Forest, XGBoost, Logistic Regression Classifier, K-Means clustering—are used to learn crime patterns and identify significant predictors. Regression models (Multiple Linear Regression, Logistic Regression, Ridge Regression) are applied to validate the index dimensions statistically. This combination ensures high predictive power, meaningful interpretability, and robustness. The validated MDCI can be used for hotspot detection, risk profiling, resource allocation, and policymaking.

V. SYSTEM ARCHITECTURE

The system architecture for the proposed hybrid ML and regression framework is designed as a multistage analytical pipeline that enables accurate validation of the Multidimensional Crime Index (MDCI) for crimes against women. The architecture begins with a comprehensive **data acquisition layer**, where structured and unstructured data are collected from police records, national crime databases, census reports, socio-economic surveys, and geospatial information systems. Once the data is aggregated, it passes through the **data preprocessing module**, which handles missing values, performs normalization, encodes categorical features, removes outliers, and constructs engineered variables essential for representing socio-economic and demographic dimensions. Following preprocessing, the **index construction layer** generates the MDCI by computing five major dimensions: crime severity, socio-economic vulnerability, demographic risk, spatiotemporal crime density, and reporting discrepancy. These dimensions together form a unified index that captures the multidimensional nature of crimes against women.

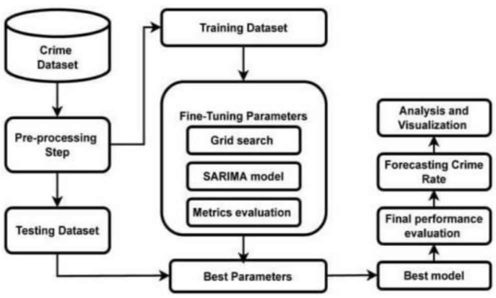

Fig.5.1: System architecture

The next stage is the **machine learning module**, in which algorithms such as Random Forest, SVM, XGBoost, and clustering techniques analyze patterns in crime data, identify dominant predictors, and generate severity or risk scores. This module enables the model to learn nonlinear relationships that traditional statistical methods may overlook. After ML pattern discovery, the system transitions into the **regression validation layer**, where statistical models—including multiple linear regression, logistic regression, and ridge regression—are applied to validate the importance, consistency, and reliability of each index dimension. This hybrid integration ensures that while ML captures complex patterns, regression methods provide interpretability and statistical confirmation. After both modules generate their outputs, the **integration and scoring layer** combines machine learning predictions with regression-based significance values to produce the final validated MDCI. The architecture concludes with a **visualization and reporting interface** that displays interactive dashboards, heat maps, risk profiles, and prediction graphs, enabling policymakers, law enforcement agencies, and NGOs to easily interpret and act upon the insights generated by the system. Overall, the architecture ensures accuracy, transparency, and practical usability through the seamless fusion of ML-driven intelligence and regression-based validation.

VI. IMPLEMENTATION

Fig.6.1: Home page

Fig.6.2.: Service provider login page

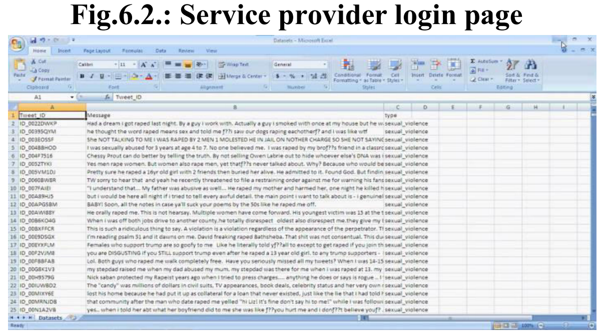

Fig.6.3: Dataset

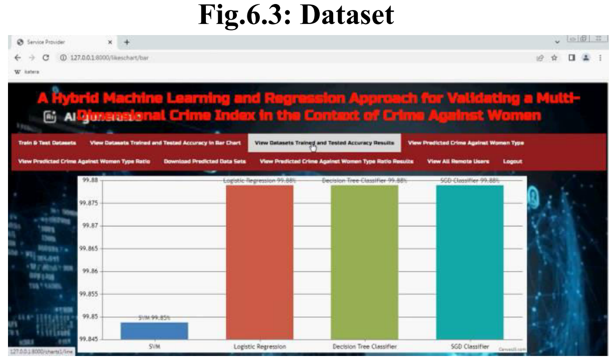

Fig.6.4: Algorithms and accuracy

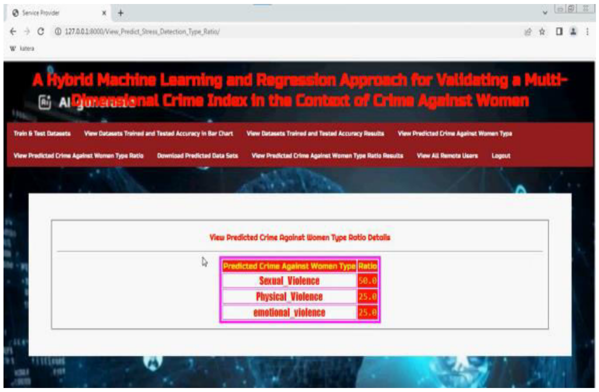

Fig.6.5: Predicted crime against women type ratio

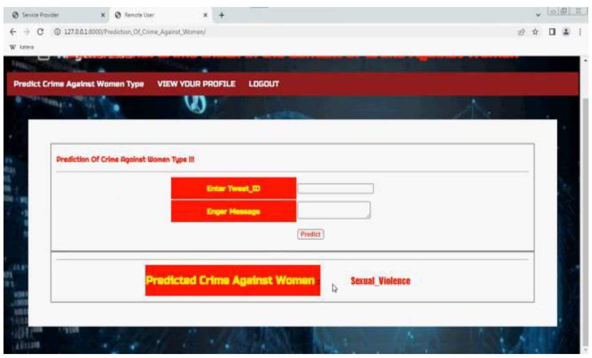

Fig.6.6: Predicted output

VII. CONCLUSION

The research presents an integrated framework for validating a Multidimensional Crime Index for crimes against women using a hybrid ML and regression approach. This dual-method strategy enhances both predictive accuracy and interpretability, overcoming the limitations of conventional analytical tools. The hybrid model demonstrates improved robustness when working with complex socio-economic and demographic datasets. By delivering a validated and explainable index, the system can support policymakers, law enforcement agencies, and social organizations in implementing targeted interventions, improving resource allocation, and strengthening preventive strategies. This hybrid analytical methodology contributes to the development of evidence-based and technologically informed solutions for enhancing women's safety.

VII. FUTURE SCOPE

The future scope of the proposed hybrid ML and regression framework for validating a Multidimensional Crime Index (MDCI) is vast, offering multiple directions for technological enhancement, social integration, and policy-level adoption. One promising expansion is the integration of **real-time data streams** from emergency helplines, social media platforms, public surveillance networks, and IoT-enabled smart city infrastructure, allowing the index to transition from a static analytical tool to a **dynamic, continuously updating risk prediction engine**. This would enable authorities to detect rising threats instantly and deploy preventive measures more efficiently. Furthermore, the MDCI can be strengthened by incorporating **deep learning models**, such as LSTM networks for temporal crime forecasting and transformer-based models for analyzing textual complaint data, thus improving the system's ability to interpret complex and unstructured information. Another area of advancement involves enhancing **geospatial intelligence** by integrating high-resolution satellite data, GPS mobility patterns, and GIS-based spatial clustering to produce micro-level hotspot predictions for urban and rural areas.

On the socio-behavioral front, future work can incorporate **psychological, cultural, and community-level indicators**, such as gender perception surveys, domestic conflict patterns, and historical socio-cultural biases, which influence crimes against women but remain underrepresented in traditional datasets. Additionally, incorporating **explainable AI (XAI)** techniques will be essential to ensure that the MDCI remains transparent, trustworthy, and legally usable, enabling law enforcement and policymakers to understand why certain regions are classified as high risk. The framework can also be expanded into a **multi-agency collaborative platform**, enabling NGOs, local governments, women's welfare organizations, and law enforcement agencies to share data, report incidents, and monitor interventions collectively. On the deployment side, the MDCI can be implemented in **mobile safety apps**, providing real-time alerts and risk assessments to women in vulnerable areas, while policymakers can use it to simulate the impact of laws, awareness campaigns, and policing strategies.

Internationally, the model can be adapted for **global comparative studies**, allowing researchers to analyze how socio-economic, cultural, and legal differences influence crimes against women across countries. Over time, the system may evolve into a **predictive governance tool** that not only forecasts crime but also evaluates the long-term effectiveness of interventions, identifies emerging social vulnerabilities, and guides budget allocation for safety measures. With continuous advancements in AI ethics, privacy preservation, and federated learning, the MDCI can be made fully compliant with global privacy regulations, making it scalable for nationwide and global deployment. Thus, the future scope emphasizes transforming the model into a robust, ethical, data-driven ecosystem that plays a critical role in improving women's safety, empowering vulnerable communities, and influencing evidence-based policymaking worldwide.

IX. REFERENCES

[1] K. Sukhija, S. N. Singh, and J. Kumar, ‘‘Spatial visualization approach for detecting criminal hotspots: An analysis of total cognizable crimes in the state of haryana,’’ in 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), pp. 1060–1066, IEEE, 2017.

[2] Anjali and B. Kumar, ‘‘Spatial analysis of multivariate factors influencing suicide hotspots in urban tamil nadu,’’ Journal of Affective Disorders Reports, vol. 16, p. 100741, 2024.

[3] Á. González-Prieto, A. Brú, J. C. Nuño, and J. L. González-Álvarez, ‘‘Hybrid machine learning methods for risk assessment in gender-based crime,’’ Knowledge-Based Systems, vol. 260, p. 110130, 2023.

[4] G. V. Manish, Simran, J. Kumar, and D. K. Choubey, ‘‘Identification of hotspot of rape cases in nct of delhi: A data science perspective,’’ in International Conference on Information Systems and Management Science, pp. 485–496, Springer, 2021.

[5] G. Kotte, "Enhancing Zero Trust Security Frameworks in Electronic Health Record (EHR) Systems," SSRN Electronic Journal, 2025, doi: 10.2139/ssrn.5283668.

[6] V. Ceccato and A. Loukaitou-Sideris, ‘‘Fear of sexual harassment and its impact on safety perceptions in transit environments: a global perspective,’’ Violence against women, vol. 28, no. 1, pp. 26–48, 2022.

[7] C. M. Spencer, S. M. Stith, and B. Cafferky, ‘‘What puts individuals at risk for physical intimate partner violence perpetration? a meta-analysis examining risk markers for men and women,’’ Trauma, Violence, & Abuse, vol. 23, no. 1, pp. 36–51, 2022.

[8] G. Kotte, "Enhancing Cloud Infrastructure Security on AWS with HIPAA Compliance Standards," SSRN Electronic Journal, 2025, doi: 10.2139/ssrn.5283660.

[9] P. K. Saravag and B. R. Kumar, ‘‘An application of scan statistics in identification and analysis of hotspot of crime against women in rajasthan, india,’’ Applied Spatial Analysis and Policy, 2024.

[10] M. Flood and B. Pease, ‘‘Factors influencing attitudes to violence against women,’’ Trauma, violence, & abuse, vol. 10, no. 2, pp. 125–142, 2009.

[11] S. T. Reddy Kandula, "Comparison and Performance Assessment of Intelligent ML Models for Forecasting Cardiovascular Disease Risks in Healthcare," 2025 International Conference on Sensors and Related Networks

(SENNET) Special Focus on Digital Healthcare(64220), pp. 1–6, Jul. 2025, doi: 10.1109/sennet64220.2025.11136005.

[12] S. N. Ogden, M. E. Dichter, and A. R. Bazzi, ‘‘Intimate partner violence as a predictor of substance use outcomes among women: A systematic review,’’ Addictive behaviors, vol. 127, p. 107214, 2022.

[13] S. Srivastava, P. Kumar, Rashmi, R. Paul, and P. Dhillon, ‘‘Does substance use by family members and community affect the substance use among adolescent boys? evidence from udaya study, india,’’ BMC Public Health, vol. 21, pp. 1–10, 2021.

[14] Todupunuri, A. (2022). Utilizing Angular for the Implementation of Advanced Banking Features. Available at SSRN 5283395.

[15] K. Shorette and R. Burroway, ‘‘Consistencies and contradictions: Revisiting the relationship between women’s education and infant mortality from a distributional perspective,’’ Social Science Research, vol. 105, p. 102697, 2022.

[16] J. Chopin, E. Beauregard, and N. Deslauriers-Varin, ‘‘Less exposed, more vulnerable? understanding the sexual victimization of women with disabilities under the lens of victimological theories,’’ International Review of Victimology, p. 02697580231154941, 2023.

[17] Z. Rodriguez, ‘‘The power of employment: Effects of india’s employment guarantee on women empowerment,’’ World development, vol. 152, p. 105803, 2022.

[18] T. A. Craigie, V. Taraz, and M. Zapryanova, ‘‘Temperature and convictions: Evidence from india,’’ Environment and Development Economics, 2023.