Spatial Patterns of Violence Against Women and Children using Geographic Information System and Density-Based Clustering Algorithm

Roger S. Mission [1]
[1]College of Computer Studies, University of Antique, Sibalom, Antique, Philippines
rsmission@antiquespride.edu.ph

***Abstract*—This paper underscores the importance of the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm in data mining. In this research, data mining clustering methods were applied to investigate consummated felonies related to the ”Anti-Violence Against Women and Their Children Act of 2004 - RA 9262” from 2018 to 2023. The criminal data processed from the Police Regional Office 6 of the Philippine National Police encompassed 248 recorded cases reflecting cases over the specified period. The significance of this study lies in its utilization of ArcGIS Pro software to process the provided data through clustering techniques, presenting a robust approach for detecting criminal activities and recognizing patterns to aid law enforcement in crime reduction efforts. Spatial data mining proves practical when dealing with geographic crime datasets, facilitating the analysis of large volumes of crime data. The DBSCAN algorithm was employed to cluster crime incidents centered on predefined events, with the resultant clusters used to identify hotspots. These clustering outcomes are then visualized using GIS, enabling real-time mapping of crime distribution for law enforcement agencies to comprehend and engage with effectively. The outcomes empower stakeholders to devise interventions tailored to specific locations, thereby contributing to a safer environment for women and children. The study illuminates the localized analysis of crime distribution, offering insights into the interconnected factors influencing criminal incidents and providing a framework for crafting targeted and efficient strategies for crime prevention, thereby enriching the broader dialogue on crime management and public safety.**

***Index Terms*—DBSCAN algorithm, VAWC, Crime clustering, GIS, spatial, Hotspots**

I. Introduction

In response to the growing issue of crime, criminology increasingly relies on advanced research methods to collect and analyze crime data. The rise in criminal activities emphasizes the need for law enforcement to proactively manage and reduce unlawful behavior. This study merges Information Technology (IT) with the field of criminological research. This integration allows for the use of advanced techniques like clustering and pattern recognition to analyze criminal activity data.. The focus is on predictive hotspot mapping to identify and forecast potential crime locations, enhancing data collection, storage, and analysis for a more comprehensive approach to understanding and managing crime. The use of IT in this study goes beyond improving data management, playing a crucial role in real-time crime monitoring, information sharing among law enforcement agencies, and developing predictive models. This integration empowers law enforcement to not only identify crime hotspots but also to respond quickly to emerging trends, allocate resources strategically, and implement targeted interventions. As technology advances, the synergy between criminology, clustering techniques, pattern recognition, and information technology becomes increasingly important for a proactive, data-driven approach to addressing crime complexities. Advances in survey and remote sensing technology have greatly increased the ability to gather large amounts of geographic data in the past decade. However, synthesizing and interpreting this data intelligently is crucial. Geographic Information Systems (GIS) are essential tools for turning geographic data into insightful knowledge. GIS helps analyze geographical occurrences connected to crime, transforming information technology, criminology, medicine, and other professions. GIS enables the display, investigation, and interpretation of spatial data, offering a new perspective on understanding criminal behavior. Integrating data from various sources can help create comprehensive spatial models with high accuracy in predicting future hotspots and revealing crime patterns. The study recognizes the value of hotspot maps as effective policing tools for understanding areas and potential causes of crime. To address this, the research aims to go beyond statistics by using GIS and Density-Based Clustering Algorithm to analyze Violence Against Women and Children (VAWC) data, revealing spatial patterns of this complex issue. This knowledge can empower stakeholders to design location-specific interventions, contributing to a safer environment for women and children.

II. OVERVIEW OF RELEVANT LITERATURE

The exploration of spatial patterns in violence against women and children (VAWC) has gained traction in recent years, with Geographic Information Systems (GIS) and density-based clustering algorithms playing a key role. Studies by [1], [2] showcase the effectiveness of GIS in identifying hotspots of VAWC, highlighting areas with high incident rates. Kernel Density Estimation (KDE), a common density-based algorithm, is employed in [3] to visualize these hotspots, enabling targeted resource allocation and prevention strategies. Furthermore, [4] demonstrates the utility of DBSCAN (Density-Based Spatial Clustering of Applications with Noise) in uncovering hidden clusters of VAWC incidents, potentially

ISSN:3006-2985(Print)
ISSN:3006-9297(Online)

revealing underlying risk factors associated with specific locations. This spatial analysis approach offers valuable insights for policymakers and NGOs working to combat VAWC, allowing for data-driven interventions and improved resource allocation.

III. THEORETICAL FRAMEWORK

A. Basic Theory of GIS Spatial Patterns

A core principle of Geographic Information Systems (GIS) lies in the analysis of spatial patterns. This theory posits that the distribution of features and events across Earth's surface isn't random, but rather exhibits meaningful clusters that reveal deeper processes, relationships, and trends. Key concepts like clustering, dispersion, and randomness form the foundation for this analysis. By examining how features are arranged in relation to one another – considering factors like proximity, density, connectivity, and orientation – GIS unveils these spatial patterns [5]. Identifying such patterns is crucial for informed decision-making across various fields. This research will examine into the spatial distribution patterns of violence against women and children, aiming to optimize resource allocation, identify areas where intervention is most needed, and potentially predict future trends in these incidents. The application of Geographic Information Systems (GIS) in crime analysis rests upon the principles of spatial analysis and powerful database integration. By visualizing crime incidents spatially, GIS enables analysts to identify crime hotspots, where offenses cluster, facilitating targeted interventions. Furthermore, GIS allows for the combination of crime data with socioeconomic variables, infrastructure maps, and environmental factors, providing a rich context for understanding crime patterns. This has several advantages: hotspots become visually clear, supporting resource optimization; spatiotemporal patterns emerge, helping predict future crime risks; and GIS-generated outputs enhance data-driven strategic decision-making in law enforcement. However, limitations exist. GIS results are contingent on the quality of input data – incompleteness or bias can skew analysis. Geocoding errors may misrepresent crime locations. Finally, technical expertise is required, as specialized training is necessary to fully leverage the potential of GIS.

B. Density-based Clustering

The Density-based clustering is a crucial GIS technique for identifying areas where features are highly concentrated, known as 'clusters'. Unlike partitioning methods like K-means, density-based algorithms don't require a predetermined number of clusters. Instead, they discover clusters based on the density of points within a specified search radius [6]. Points within dense regions form clusters, while points in low-density areas are classified as noise or outliers [6], [7]. This method is particularly useful in geographic analysis because it can detect clusters of varying shapes and sizes, revealing complex spatial patterns. Density-based algorithms, such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise), are widely applied in GIS for tasks ranging from crime hotspot analysis to disease outbreak mapping [8]. DBSCAN (Density-Based Spatial Clustering of Applications with Noise), introduced by Ester in 1996 is a foundational density-based clustering algorithm. Its core idea centers on identifying regions of high density in the data, separated by regions of lower density. Two key parameters, Eps and MinPts, are defined: Eps specifies the radius of a neighborhood around a data point, and MinPts sets the minimum number of points required within that neighborhood to be considered a "core point". The algorithm begins with an arbitrary starting point. If it's a core point, all points within its Eps-neighborhood are retrieved. Iteratively, newly discovered core points expand the cluster. Points not within a dense neighborhood are labeled as noise. This approach empowers DBSCAN to discover clusters of arbitrary shapes in the dataset while effectively isolating outlier points as noise, a significant advantage over many traditional clustering methods.

IV. METHODOLOGY/RESEARCH DESIGN

A. Data

The crime dataset comes from the Police Regional Office (PRO) 6 of the Philippine National Police. It was requested via the foi.gov.ph website. This dataset presents data on the consummated felony with 248 recorded cases related to the "Anti-Violence Against Women and Their Children Act of 2004 - RA 9262" for the years 2018 to 2023 in the Province of Antique, Philippines. This Anti-Violence Against Women and Their Children Act of 2004 (Republic Act No. 9262) is a Philippine law enacted to define violence against women and their children, provide protective measures for victims, and prescribe penalties for perpetrators [9]. The geographic coordinates of Antique is located at longitude 122.08333000 and latitude 11.16667000. The GPS coordinates for the Province of Antique are 11° 10' 0.012" N and 122° 4' 59.988" E [10]. The Province of Antique landed in 61st place out of 82 Philippine provinces in the 2023 Cities and Municipalities Competitiveness Index (CMCI). This ranking is based on the combined scores of each province's cities and municipalities across four key areas: economic dynamism, government efficiency, infrastructure, and resiliency [11]. It is divided into 18 municipalities and 420 barangays. The major industries in Antique are agriculture, fishing, and tourism. The study, which focused on spatial analysis, looked at crucial aspects of crime data: the location of crimes, date committed, their description, latitude and longitude as shown in Table I.

B. Density-Based Spatial Clustering of Applications with Noise

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a powerful unsupervised clustering algorithm used widely in spatial data analysis and proposed by Ester in 1996. DBSCAN doesn't require a pre-determined number of clusters. Instead, it focuses on identifying regions of high density in a dataset, classifying points in low-density regions as noise or outliers [7]. This approach makes it particularly robust in handling datasets with noise and discovering

TABLE I
RAW SPREADSHEET CONTAINING CRIME DATASET

DATE COMMITTED	TIME COMMITTED	STAGES OF FELONY	LATITUDE	LONGITUDE
2023-04-30	08:25:00	CONSUMMATED	10.50022752	121.9279062
2023-05-12	23:30:00	CONSUMMATED	10.59735564	121.974444
2023-11-13	21:20:00	CONSUMMATED	10.68348066	121.9937624
2023-04-04	22:30:00	CONSUMMATED	10.7035089	121.9737111
2023-06-26	03:10:00	CONSUMMATED	10.7035089	121.9737111
2023-08-31	09:00:00	CONSUMMATED	10.73370482	121.9550128
2023-03-22	10:45:00	CONSUMMATED	10.74163382	121.9501738
2023-03-08	17:30:00	CONSUMMATED	10.78094721	122.0069695
2023-01-15	20:30:00	CONSUMMATED	10.80295182	121.9555636
2023-01-29	18:00:00	CONSUMMATED	10.80295182	121.9555636
2023-01-20	22:40:00	CONSUMMATED	10.80295182	121.9555636
2023-09-25	01:00:00	CONSUMMATED	10.98254558	122.0787961
2023-03-11	07:25:00	CONSUMMATED	11.01214026	122.1359489

clusters of complex and irregular shapes. DBSCAN is widely used in various fields, including GIS, for tasks like identifying crime hotspots, analyzing disease outbreaks, and detecting patterns in customer behavior [12]. The key to the DBSCAN clustering algorithm is the determination of the two parameters Eps and MinPts, where Eps represents the neighborhood distance threshold, and MinPts represents the critical value of the number of samples in the neighborhood [12]. As Figure 1,if there is a data set $D = X_1, X_2, ..., X_n$. The basic definition involved in DBSCAN is as follows [13]: Core object. For any sample $X_j \in D$, if it contains at least MinPts samples in its Eps neighborhood, then X_j is called the core object. Direct density. If X_i is located in the Eps neighborhood of X_j, and X_j is the core object, then X_i is directly reached by the density of X_j. The density is reachable. If there is a data set sequence $p_1, p_2, ..., p_T$ that satisfies $p_1 = X_i, p_T = X_j$, and p_{T+1} is directly reached by the density of p_T, then it is said that X_j is reachable by the density of X_i, that is, the reachability of the density is transitive. Definition 4: Density is connected. If there is a core object sample X_k, both X_i and X_j can be reached by the density of X_k, and the density of X_i and X_j is said to be connected. As shown in the figure below, X_1 and X_2 are the core points, X_3 and X_4 are the boundary points, X_2 is directly reached by the density of X_1, X_3 is directly reached by the density of X_2, X_4 is reachable by the density of X_1, and the density of X_4 and X_3 are connected. The fundamental

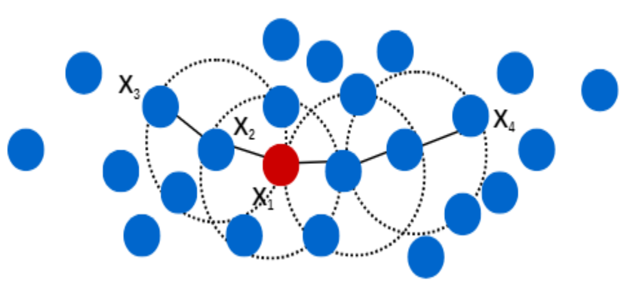


Fig. 1. Schematic diagram of DBSCAN

idea of DBSCAN clustering is to start with a random data point and check if it has enough neighbors within a defined radius to be considered a core point. If it does, DBSCAN gathers all points directly connected to this core point, forming the initial cluster. Then, it iteratively expands the cluster by including any points directly connected to the existing cluster members. This process continues until no more points can be added, completing one cluster. If the starting point isn't a core point, it's temporarily labeled as noise. This process repeats, picking new unprocessed points until all data points have been clustered or marked as noise. Table 2 shows the flowchart to visualize and analyze the spatial distribution of crime against women and their children.

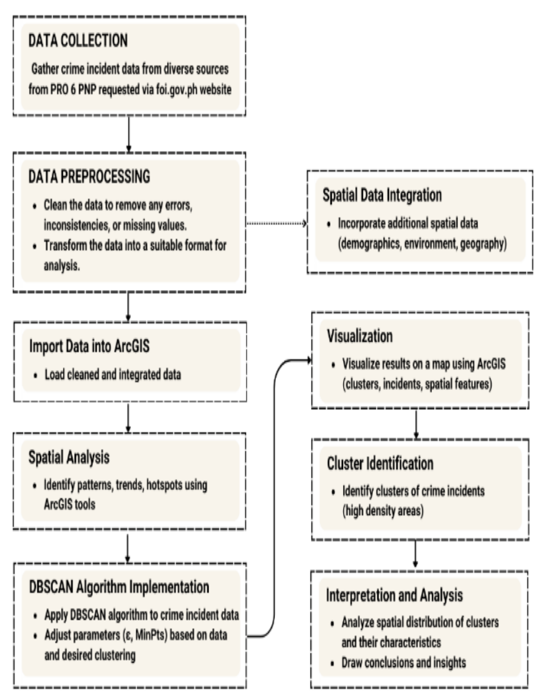


Fig. 2. Flowchart to visualize and analyze the spatial distribution of crime against women and their children

C. Spatial Statistical Methods

This study used ArcGIS Pro (version 3.2.2) to analyze the spatial patterns of consummated felonies. The researcher calculated the mean center to find the average location of crimes and used choropleth maps to visualize their distribution across different areas. Spatial autocorrelation (measured by Moran's Index) was used to understand if similar crime incidents tended to happen near each other. A positive Moran's Index means crimes cluster in specific locations, while a negative index suggests they are scattered. A value near zero means the distribution is likely random. In this analysis, the term "cluster" indicates a group of crime incidents that occur significantly close together, either in terms of geography, time, or both.The Moran's Index statistic for spatial autocorrelation is given as:

$$I = \frac{n \ \sum_{i=1}^{n} \cdot \sum_{j=1}^{n} w_{i,j} z_i z_j.}{S_0 \ \sum_{i=1}^{n} z_i^2.}$$

where z_i is the deviation of an attribute for feature i from its mean $(x_i - \bar{x})$, is the spatial weight between feature i and j, n is equal to the total number of features, S_0 is the aggregate of all the spatial weights:

$$S_0 = \sum_{i=1}^{n} \cdot \sum_{j=1}^{n} w_{i,j}.$$

The z_I score for the statistic is computed as:

$$z_I = \frac{I - E[I]}{\sqrt{V[I]}}$$

where:

$$E[I] = -\frac{1}{n-1}$$

$$V[I] = E\left[I^2\right] - E\left[I^2\right]$$

V. PRESENTATION AND DISCUSSION OF RESULTS

A. Spatial Patterns of Violence Against Women and Children

The Philippines has made important progress in combating violence against women and children (VAWC) with the Anti-Violence Against Women and Their Children Act of 2004 (RA 9262). This law identifies diverse forms of abuse (physical, sexual, psychological, and economic), offering a system of protection, legal action, and support for victims. RA 9262 has increased awareness of VAWC, enhanced responses, and encouraged victims to come forward. It includes measures like protection orders, a VAWC hotline, and local VAWC desks. Research indicates the law has boosted VAWC reporting and improved victim services. Despite this, hurdles remain, such as limited awareness in some areas, underfunded support services, and attitudes blaming victims. Overcoming these obstacles demands ongoing cooperation between the government, civil society, and the public to fully implement the law's goals and build a society where women and children are safe. The provided data in Table 2 shows the number of Violence Against Women and Children (VAWC) cases reported in a certain area over the years 2018 to 2023.

TABLE II
CRIME DATASET CASES REGISTERED DURING THE 2018-2023

Stages of Felony	Offense	Year	Total
Consummated	Violence Against Women and Children Act of 2004-RA 9262	2018	21
		2019	38
		2020	34
		2021	42
		2022	56
		2023	57
TOTAL			248

Table II presents data on registered VAWC (Violence Against Women and Children) cases categorized as consummated felonies over the period 2018–2023. The data reveals a disturbing upward trend in the number of registered cases. From 2018 to 2022, there is a steady increase in VAWC cases annually, with a significant jump between 2021 (42 cases) and 2022 (56 cases). The total for 2023 was 57 cases. The dataset highlights a total of 248 registered VAWC cases across the six-year period, emphasizing the persistent and increasing nature of this serious issue. Using DBSCAN algorithm, spatial distribution of recorded VAWC cases in 2018-2023 as shown in Figure 345678. Across the years 2018-2023, the spatial analysis consistently revealed variations in the density of reported VAWC cases. Distinct clusters emerged, often demonstrating areas of high VAWC incidence alongside those with lower or more scattered occurrences. In certain years (2018, 2021, 2022), a single cluster of high-density cases was identified, while other years (2019, 2020) showed two distinct clusters. The presence of "noise" (Cluster -1) highlights isolated incidents or areas where data may be inconsistent or insufficient for clear clustering. These findings underscore the complex spatial nature of VAWC, suggesting some areas experience persistently higher levels of violence while others have more sporadic patterns. This information is crucial for targeted interventions and resource allocation in the fight against violence against women and children.

B. Spatial autocorrelation of VAWC using Moran's Index statistic

The Moran's Index statistic for spatial autocorrelation is shown in the Figure 9 below. Figure 9 shows the spatial autocorrelation of VAWC cases using Moran's Index of -0.011915. This indicated a very slight tendency toward spatial dispersion in the recorded cases of VAWC between 2018 and 2023. This means that there was a minimal pattern where areas with high numbers of VAWC cases tended to be located near areas with low numbers of cases, and vice versa. The analysis of the dataset employed Euclidean distance, the most common and intuitive way to measure the straight-line distance between data points. In this case, it likely calculated distances between VAWC case locations. A distance threshold of 15399.1136 meters was applied, indicating a focus on identifying spatial relationships (such as clustering or dispersion) between VAWC cases occurring within this radius of each other. This approach aimed to reveal whether VAWC cases exhibited spatial patterns that deviated from what would be expected by random chance. However, this pattern was not statistically significant. The z-

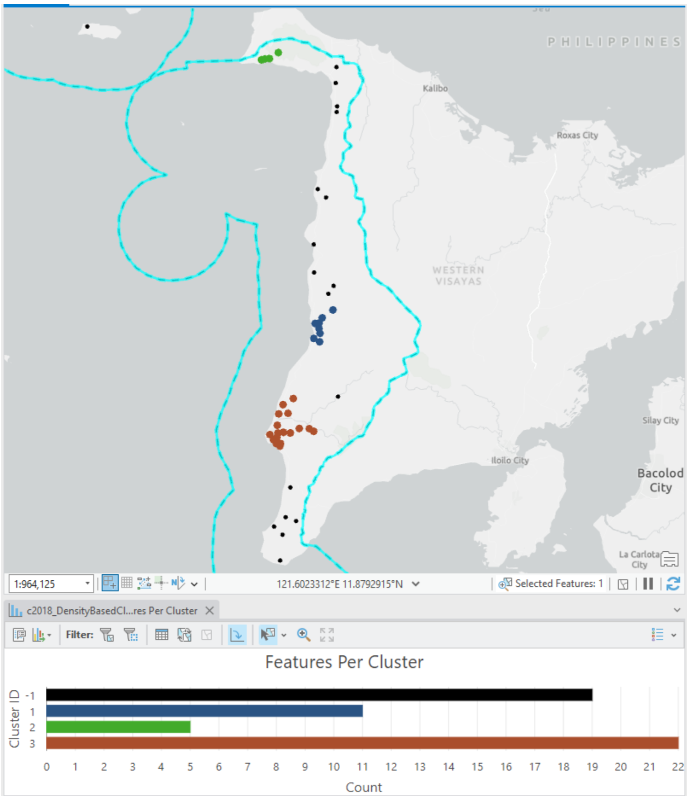


Fig. 3. Spatial distribution of recorded VAWC cases in 2018 using ArcGIS and DBSCAN algorithm

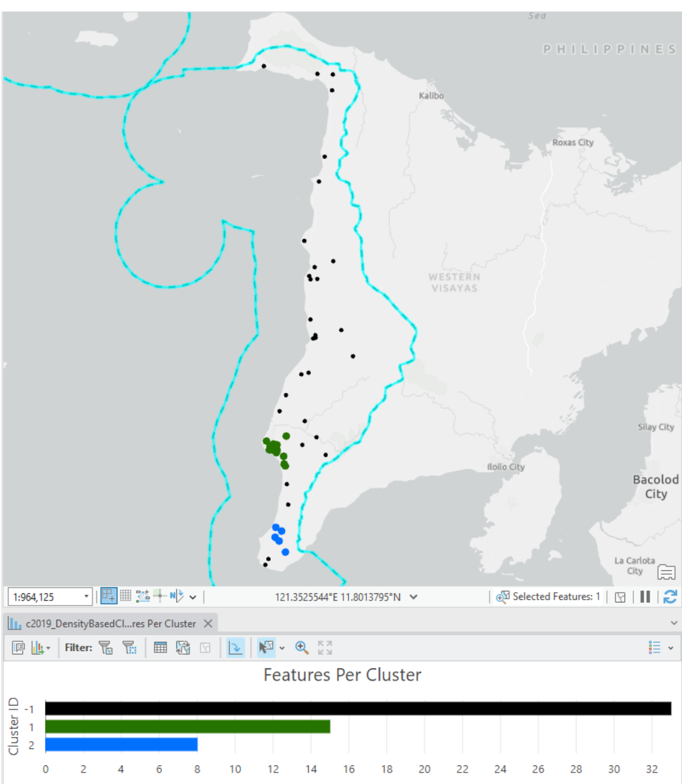


Fig. 4. Spatial distribution of recorded VAWC cases in 2019 using ArcGIS and DBSCAN algorithm

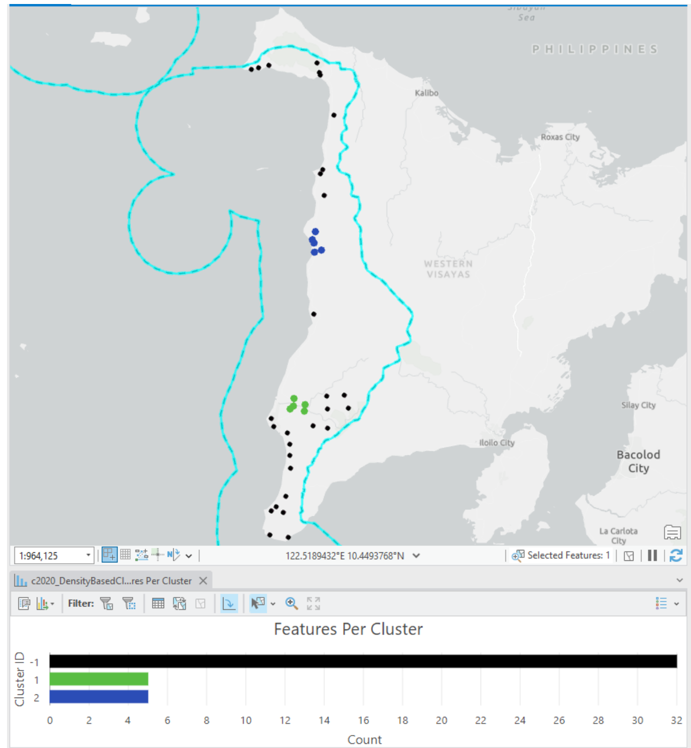


Fig. 5. Spatial distribution of recorded VAWC cases in 2020 using ArcGIS and DBSCAN algorithm

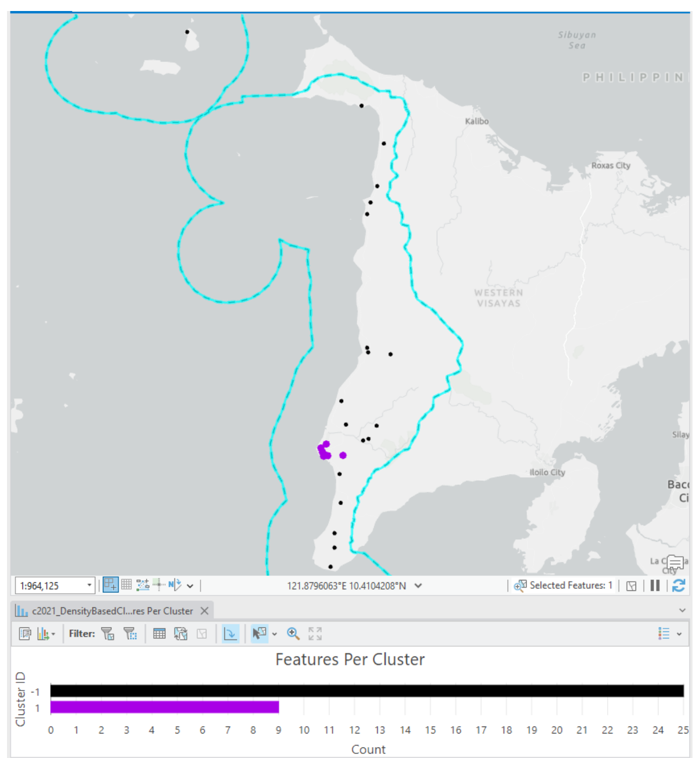


Fig. 6. Spatial distribution of recorded VAWC cases in 2021 using ArcGIS and DBSCAN algorithm

score of -0.457886 and the associated p-value of 0.647034 demonstrated that the observed spatial pattern could easily

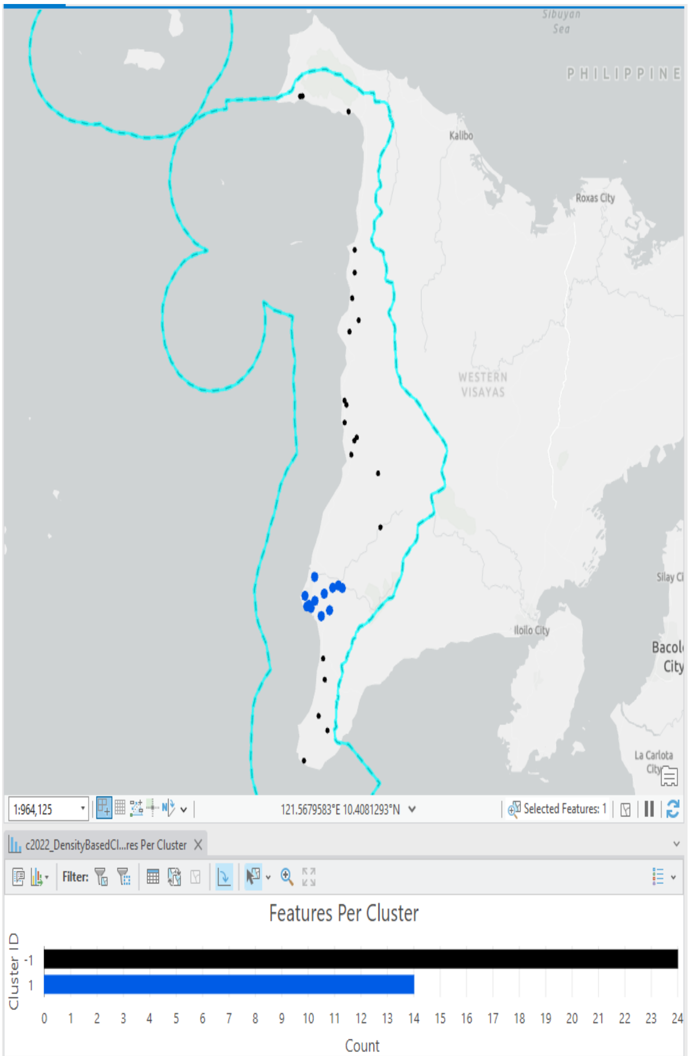


Fig. 7. Spatial distribution of recorded VAWC cases in 2022 using ArcGIS and DBSCAN algorithm

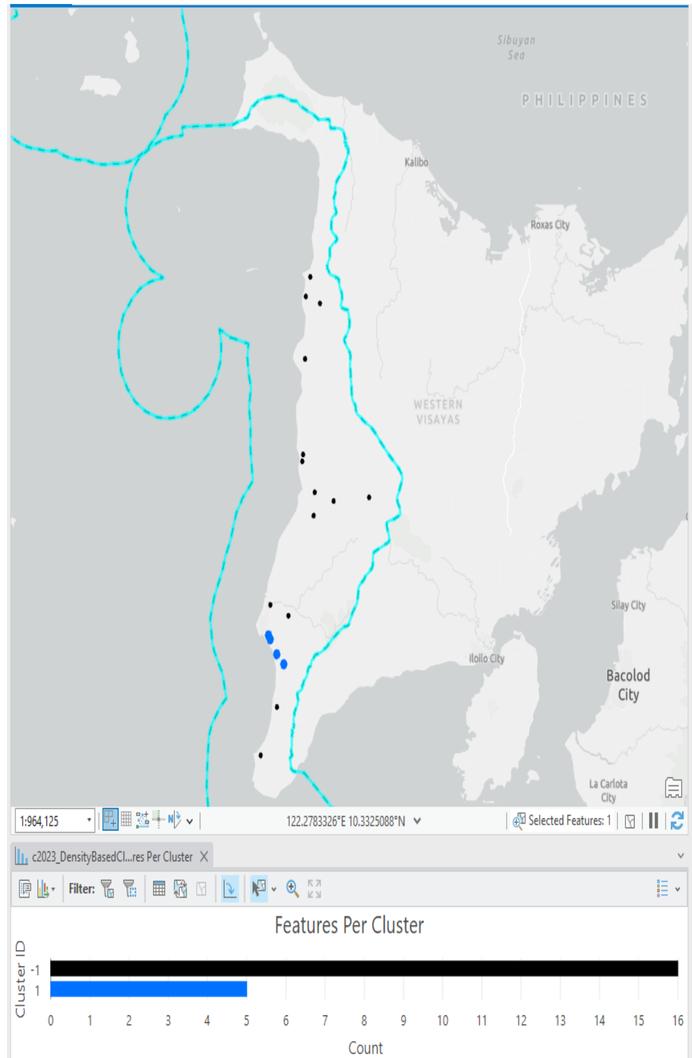
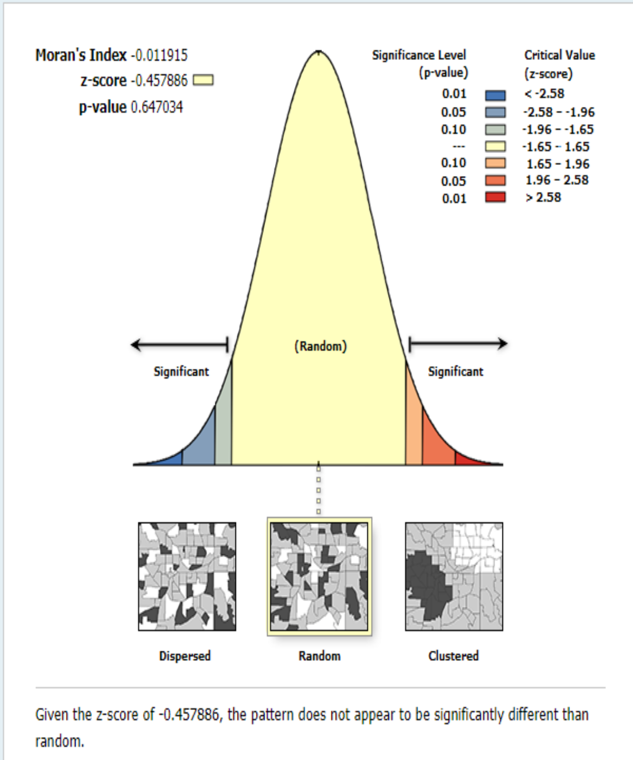


Fig. 8. Spatial distribution of recorded VAWC cases in 2023 using ArcGIS and DBSCAN algorithm

have occurred by random chance. Several factors could contribute to a pattern of spatial dispersion in VAWC cases. Underreporting might vary between regions, with areas offering strong support systems, increased awareness of VAWC, or reduced social stigma potentially seeing higher reported cases compared to neighboring areas with less support. Uneven distribution of resources such as shelters, counseling centers, and legal aid could also have an impact on reporting rates, as victims with limited access to these services may be less likely to come forward. Additionally, socioeconomic disparities between areas, such as poverty, unemployment, and education levels, can influence both the occurrence and reporting of VAWC. Moreover, there is no strong evidence to suggest any meaningful clustering or dispersion of VAWC cases within the analyzed region.

VI. CONCLUSIONS

This study highlights the potential of integrating spatial analysis, information technology, and criminological principles to better understand and address critical issues like Violence Against Women and Children (VAWC). By using clustering algorithms and GIS tools, the researcher moves beyond simplistic crime statistics to visualize hotspots, reveal complex spatial patterns, and potentially anticipate areas where future incidents might occur. The observed fluctuations in VAWC reporting emphasize the need for ongoing monitoring and may reflect the impact of awareness campaigns and evolving reporting systems. The spatial analysis reveals a complex landscape with variations in VAWC density. Identifying clusters of high violence emphasizes the need for targeted interventions within these hotspots. Additionally, the presence of "noise"—seemingly isolated cases—requires careful examination. These sporadic instances could be due to underreporting or reflect unique social factors requiring tailored solutions. Although the Moran's Index in this specific case didn't indicate statistically significant clustering, its use showcases the value of investigating spatial autocorrelation in crime data. This research offers opportunities to enhance law enforcement strategies. Understanding VAWC's spatial dynamics facilitates more effective resource allocation. By pinpointing areas at risk for future



Global Moran's I Summary

Moran's Index	-0.011915
Expected Index	-0.004049
Variance	0.000295
z-score	-0.457886
p-value	0.647034

Fig. 9. Spatial autocorrelation of VAWC Cases using Moran's Index

incidents and facilitating preventative measures, predictive hotspot mapping enables proactive policing. Moreover, IT integration is key: streamlined data collection, real-time information sharing, and sophisticated crime analysis tools all enhance law enforcement capabilities. The study acknowledges the importance of addressing VAWC's specific issue. The Philippines' Anti-Violence Against Women and Their Children Act of 2004 (RA 9262) plays a significant role in raising awareness and strengthening victim support. Yet, challenges persist. Spatial analysis provides an additional tool to bolster this law's implementation. Identifying VAWC hotspots allows for targeted action like community education, tailored victim services, and focused law enforcement. Finally, this interdisciplinary approach combining spatial analysis, information technology, and criminology offers a robust framework for understanding and combating crime. As technology evolves, this integrated toolkit will become ever more essential for data-driven policing, proactive crime prevention, and ultimately, the creation of safer communities.

Acknowledgment

I would like to express my sincere gratitude to the Freedom of Information (FOI) website (foi.gov.ph) and the Philippine National Police (PNP) Region 6 for their invaluable support throughout this research. The comprehensive dataset they provided was fundamental in conducting the spatial analysis and gaining insights into crime patterns. I am deeply appreciative of their commitment to transparency and their efforts to make data accessible for research that can contribute to safer communities. I would also like to extend my thanks to the individuals who shared their expertise and guidance throughout this research endeavor.

REFERENCES

[1] A. O. Dogru, D. R. Magano, N. Ulugtekin, C. Goksel, D. Z. Seker, and S. Sozen, "A GIS-based spatiotemporal analysis of violent trauma hotspots in Vancouver, Canada: identification, contextualisation and intervention," BMJ open, vol. 4, no. 2, p. e003624, 2014.

[2] K. Muldoon, L. P. Galway, A. Reeves, and K. Sampsel, "Spatial analyses using ArcGIS to visualize the distribution of SGBV cases across the Ottawa-Gatineau area in 2015," ResearchGate, figshare, [Online]. Available: [invalid URL removed] [Accessed on: 2024-04-19].

[3] X. Zhang, X. Zhang, Y. Qiang, Q. Tian, and X. Tang, "A segmentation method for lung nodule image sequences based on superpixels and density-based spatial clustering of applications with noise," PloS one, vol. 12, no. 12, p. e0189822, 2017.

[4] M. Deng, Q. Liu, G. Li, and T. Cheng, "A field-theory based spatial clustering method," Journal of Remote Sensing, vol. 14, no. 4, pp. 694–709, 2010.

[5] A. Getis and J. K. Ord, The analysis of spatial point patterns, Earthscan, 2010.

[6] M. Ester, H.-P. Kriegel, J. Sander and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise", in Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, 1996, pp. 226-231.

[7] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Morgan Kaufmann, 2011.

[8] X. Zhang, C. Xu, and Y. Zhang, "Spatiotemporal clustering analysis of COVID-19 in China: A study based on a novel spatial distance", Applied Geography, vol. 137, Nov. 2021. [Online]. Available: https://doi.org/10.1016/j.apgeog.2021.102600.

[9] "Latitude and longitude of Province of Antique" latlong.info, https://latlong.info/philippines/antique/province-of-antique/ (accessed November 13, 2023).

[10] "2023 Ranking of Provinces" cmci.dti.gov.ph, https://cmci.dti.gov.ph/rankings-data.php?unit=Provinces (accessed December 30, 2023).

[11] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise", in Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, 1996, pp. 226-231.

[12] Q. J. Jiang, Z. M. Jiang, D. Tang, J. M. Zeng. "Research on the method of detecting gross errors in slope safety monitoring data based on SSA-DBSCAN [J/OL]", Journal of Yangtze River Scientific Research Institute: 1-8, 2021.

[13] J. Liu, H. Qin, Z. Liu, S. Wang, Q. Zhang, and Z. He, "A Density-Based Spatial Clustering of Application with Noise Algorithm and its Empirical Research," Highlights in Science, Engineering and Technology, vol. 7, no. 2, pp. 174-179.