European Journal of Computer Science and Information Technology, 14(8),13-35, 2026
Print ISSN: 2054-0957 (Print)
Online ISSN: 2054-0965 (Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

Silent Signals: A Comparative Analysis of Machine Learning Models in Predicting the Vulnerability of Filipino Women Against Domestic Violence

Jerica B. Agapito[1], Allaine Anne V. Flores[2] Romie C. Mabborang [3] Alvin G. Bansil [4]
Gerald P. Estoperez[5] Jennifer C. Sy [6]
College of Science, Pamantasan ng Lungsod ng Maynila (University of the City of Manila)
Manila, Philippines

doi: https://doi.org/10.37745/ejcsit.2013/vol14n81335 Published November 08, 2026

Citation: Agapito J.B., Flores A.A.V., Mabborang R.C., Bansil A.G., Estoperez G.P., Sy J.C. (2026) Silent Signals: A Comparative Analysis of Machine Learning Models in Predicting the Vulnerability of Filipino Women Against Domestic Violence, *European Journal of Computer Science and Information Technology*, 14(8),13-35

Abstract: *Domestic violence against women remains a persistent and underreported issue in the Philippines. This study applies five (5) machine learning models, namely: KNN, Logistic Regression, Random Forest, XGBoost, and LightGBM, to predict the vulnerability of Filipino women to domestic violence using data from the 2022 NDHS. These models were evaluated using comprehensive metrics, including Accuracy, Precision, Recall, F1-Score, Log Loss, MCC, ROC-AUC, PRC-AUC, and Confusion Matrix. Logistic Regression emerged as the most effective model, demonstrating a strong balance between interpretability and predictive performance, particularly in Accuracy, F1-Score, PRC-AUC, and ROC-AUC. The analysis identified key predictive variables that emerged from the best-performing model's feature importance, namely: household composition, reproductive health knowledge, unmet need for contraception, frequency of internet usage/media exposure, asset ownership, and age of household heads. Based on these findings, several policy recommendations are proposed: expanding VAW desk services and barangay-level response systems; integrating domestic violence screening into family planning and maternal health programs; designing age-specific interventions such as economic support and education; and encouraging the ethical use of machine learning tools by social welfare agencies for early detection. Ethical considerations were also examined, including data privacy, bias mitigation, and responsible AI deployment. This study contributes to the growing field of data-driven policy making by demonstrating the value of machine learning in informing targeted and preventative social interventions.*

Keywords: Machine learning, domestic violence, predictive analysis

INTRODUCTION

Sad, apathetic, lonely, angry, quarrelsome– these are some of the heartbreaking descriptions of themselves; the women who were victimized by their intimate partners [1]. Domestic Violence (DV) is defined as a trajectory of behaviors in any formed relationship that exploits the vulnerability of their intimate partners and uses such dynamics to either gain or maintain power and control [2]. For women, it may take in various forms such as but not limited to: *"acts of abusive sexual contact, forced engagement in sexual acts, attempted or completed sexual acts without her consent, sexual harassment, verbal abuse, threats, exposure, unwanted touching, incest, etc."* [3].

In the Philippines, DV remains a timely and relevant social issue and most of the victims of this unfavorable circumstance would be Filipino Women. The Philippine Statistics Authority (PSA) reported a significant number of women have experienced physical, sexual, or psychological forms of violence. PSA also recorded 17.5% Filipino women, ranging from 15 years old up to 49 years old, have undergone horrendous acts of violence [4]. Despite the implementation of the Anti-Violence Against Women and their Children Act of 2004 and other co-existing laws and crafted bills—factors such as cultural norms, societal stigma, and limited access to governmental support services hinder and prevent the victims of DV from seeking help, further assistance, and protection.

Various studies have been established in correspondence to the contributing factors of DV against women with the usage of distinct and popular regression analysis methodologies such as Logistic Binary, Multiple Linear, and Bayesian, alongside statistical analysis for additional references regarding domestic violence and its background. However, DV lacks analysis of machine learning (ML) algorithms to test its effectiveness and accuracy, especially in the local scope. With this, the main objective of this study is to delve into a recurring social issue of DV and establish a comparative analysis between machine learning algorithms. The utilization of these ML algorithms enables the researchers to better predict and comprehend the underlying risk factors associated with DV against women in the Philippines, establish a correlation or association of these factors to the risks of domestic violence, and contribute to recommendations with the findings and supported studies. The researchers garnered the data from the 2022 National Demographic Health Survey (NDHS), conducted every five (5) years by the PSA, which was collected starting from May 2 until June 22, 2022, covering a national sample of an estimated 28,000 women. This study also aims to explore the implementation of machine learning techniques for datasets primarily: data cleaning techniques, data imputation, various feature engineering methodologies, sampling techniques for data balancing, as well as the implementation of evaluation metrics suitable for the dataset.

Assessing the best performing ML algorithm in terms of appropriate and reliable evaluation metrics paves way to positive implications and establishing relevant associations in the development of a predictive analysis in the lenses of DV in the Philippines– further yielding preventive measures through the advantageous utilization of artificial intelligence (AI) as a technological tool for advancement in the field of social sciences.

A. Conceptual Framework

The following flowchart will serve as the conceptual framework of the study which follows the Input, Pre-processing and Processing, and Output mechanism.



Figure 1. *Conceptual Framework*

B. Statement of the Problem

The study seeks to determine the best-performing machine learning algorithm to identify the vulnerability of Filipino Women and to contribute recommendations in line with the findings of the study. Furthermore, this study intends to answer the following:

1. What machine learning algorithms can be used in predicting the vulnerability of Filipino Women against Domestic Violence based on theoretical findings?
2. Among the chosen machine learning algorithms, which demonstrates the best performance in identifying women vulnerable to domestic violence in the Philippines, based on evaluation metrics such as: Accuracy, Precision, Recall, F1-Score, Log Loss, MCC, ROC-AUC, PRC-AUC, and Confusion Matrix?
3. From the variables established by the Demographic Health Survey (DHS), which factors are most relevant in developing predictive models and serve as top significant risk indicators of domestic violence against women in the Philippines using feature importance of the best performing model?
4. What analysis can be associated/correlated with the derived top significant risk indicators of domestic violence from the feature importance of the best performing model?

LITERATURE REVIEW

Machine Learning as a Tool for Predictive Performance

The rise of Machine Learning (ML) over the years has shown great potential as a tool in predicting and forecasting several significant studies. The predictive performance in ML is associated with the capacity and ability of an algorithm to forecast outcomes accurately [5], such as analyzing historical data and identifying patterns [6]. Moreover, ML provides significant advantages in predictive analytics, including high accuracy [7] and improved reliability through model blending. ML models are highly customizable and offer automated pattern detection [8], and are scalable for large-scale, real-time predictions (e.g., XGBoost, LightGBM) [9-10]. Critically, ML can reduce human biases through data-driven predictions [11] and offers probabilistic insights for informed decision-making [12].

Despite being a powerful tool in technological advancement for training artificial intelligence (AI) models, the nature, characteristics, performance of the data, and the learning of algorithms are still the basis for how effective and efficient the machine learning algorithm is [13]. Such powerful tool can be applied to various industries' predictive performance [14] such as predictive maintenance [15] energy demand prediction [14, 16], education performance predictions [17], healthcare diagnosis and outcome prediction [18], and even public policy and crime prediction [19].

As a groundbreaking technological advancement in the aspects of predictive abilities, ML covers a range of libraries and frameworks designed to handle tasks from data processing to model training and evaluation [13]. As such, some of the ML algorithms are popular for predictive modeling, each with its strengths and applications:

Artificial Neural Networks (ANN). An ML algorithm which primarily consists of interconnected nodes, more commonly referred to as neurons, that are further organized into three (3) layers: the Input Layer, Hidden Layer, and the Output Layer, of which each neuron receives input signals, applies specific mathematical functions to these neurons, and then passes them to the next layer until an output is generated [20].

K-Nearest Neighbors (KNN). A simple, non-parametric, instance-based learning algorithm that classifies new data points based on the labels of the closest data points or their neighbors [21]. The ease of implementation and intuitive approach of KNN make this ML algorithm popular for tasks such as text mining, facial recognition, and other classification and recommendation systems [22]. Moreover, KNN does not rely on any assumption about the underlying data distribution and does not learn immediately from the training data, as data is memorized instead this algorithm as a *"lazy learner"* [23].

Light Gradient Boosting Machine (LightGBM). This ML algorithm uses a leaf-wise growth strategy and can handle large-scale data, making this well-suited for high-performance applications [24]. LightGBM is commonly used in classification, regression, and other predictive modeling tasks [25].

Logistic Regression (LR). A statistical analysis method used for modelling binary outcomes, such as yes/no or true/false results [26]. This technique estimates the probability of a certain event occurring by applying a sigmoid function or simply a logistic function; it leads input variables to a value of 0 and 1 [27]. Despite its simplicity, LR is powerful in scenarios where interpretability and efficiency are important, particularly in industries such as healthcare and finance [28].

Naïve Bayes. Described to belong in the category of simplest supervised learning algorithms [29]. This statistical classification algorithm is purely based on the Bayes Theorem, wherein it assumes the presence of one feature does not affect other features [30].

Random Forest (RF). A popular ML algorithm used for both classification and regression problems and is widely recognized for its strong performance and adaptability [31]. Random forest builds multiple decision trees using random subsets of the data and features, which helps the dataset to reduce overfitting and improve variance compared to individual decision trees [32].

Support Vector Machines (SVM). Known to be a supervised machine learning algorithm which classifies data through the process of discovering an optimal line or "hyperplane" which maximizes the distance between each class in a set number of dimensional space [33]. Moreover, it is commonly used for research that involves classification, regression, and/or outlier detection [34].

Xtreme Gradient Boost (XGBoost). Known for its regularization techniques and parallel processing capabilities, making it highly effective for tabular datasets, especially in structured data competitions [35]. XGBoost provides accurate predictions particularly in sales forecasting, stock market prediction, and energy demand forecasting [36].

The Act of Domestic Violence

The act of DV is not a new phenomenon; the terrifying experience of physical, emotional, psychological, and/or economic abuse can be seen throughout history, regardless of gender, age, race, sexual orientation, faith, or class [2].

Filipino women are not far-fetched from experiencing domestic violence in comparison with those born outside the archipelagic country, as colonial and Christian roots hold both protective and paternalistic ideology that places control on women's movements. Spanish and American colonization fostered male-centeredness and male-dominance through aspects of religious and legal frameworks—completely undermining women's freedom and rights [37]. The 333 years of Spanish rule and the strong establishment of Christianity imposed severe limitations on women's basic rights and societal roles, putting them to an inferior position of a subservient to both men and the church [38-39]. American colonization further subjected women to violence, including rape and physical harm during wartime, justified by narratives of suppressing rebellion [40]. These historical patterns have collectively granted men significant leverage in Philippine society, contributing to the sexual division of labor, the imposition of fear on women, and their increased dependency on men [37].

The 19th and 20th centuries paved a path of hope towards criminalizing abuse against women. This led to the implementation of anti-violence laws, women-based activism, and support programs [41]. This global progression, which empowered women's and feminist perspectives [42], translated into significant Philippine legal provisions like the Anti-Violence Against Women and Their Children Act of 2004, The Expanded Anti-Trafficking in Persons Act of 2012, and The Anti-Sexual Harassment Act of 1995 [43]. Consequently, numerous government and non-government organizations (e.g., Philippine Commission on Women, GABRIELA Women's Party) now actively work to protect and amplify women's voices.

However, despite such existing laws catering to women, soon-to-be approved or disapproved bills for their welfare and safety, and even the rise of women empowerment movements in pursuit of gender equality and promotion of human rights–

the rampancy of DV is still of great prevalence around the world, most especially to Filipino women. Different kinds of DV and its soaring implications can still be manifested through physical, emotional, psychological, and even economic extents to women of varying ages, places, and different degrees– figuratively cutting women's tongues in amplifying their voices for justice.

Implications of Domestic Violence to Women

Women are prioritized to be safe and secure in all places. However, a woman's own 'home' is where they encounter danger and in the hands of their loved ones, where these people should be the ones they can trust [44].

Over 10 million people each year take part in DV or IPV, with women aged 18 years and older [45]. Domestic violence may include threats, coercion, or arbitrary deprivation of liberty occurring in public or private life and is commonly committed by a husband or intimate partner. As per WHO, 1 in 3 women worldwide have experienced physical and/or sexual intimate partner violence in their lifetime, wherein women's physical, mental, and sexual health were impacted negatively [46]. DV against women implicates serious psychological, physical, and economic impacts, which often result in long-term health repercussions [47].

Results from the DHS working paper shows that DV or IPV prevalence in the Philippines has reduced with over 3% of women have experienced physical IPV, 1% have experienced sexual IPV, and 11% reported that they have experienced emotional IPV in the last 12 months of the year 2022 [48].

Kalokhe et al. performed a systematic review of 137 quantitative studies about DV experiences of Indian women; these studies have been published in the prior decade, 2007-2017. The literature found that 4 in 10 Indian women (when surveyed about multiple forms of abuse) report that they have experienced DV in their lifetime, and 3 in 10 reports that they have experienced DV in the past year [49]. These findings align with the WHO estimates, indicating a prevalence of 37.7% (95% CI: 30.9%, 43.1%) in South-East Asia, including countries such as India, Maldives, Sri Lanka, Thailand, Bangladesh, and Timor-Leste. This prevalence is notably higher compared to the WHO's regional estimates for Europe, the Western Pacific, and the Americas. Moreover, mental, physical, sexual, and reproductive health have big implications for Indian women being DV as the primary reason.

In Ethiopia, a systematic review performed by Semahegn & Mengistie revealed the results that range from 20% to 78% among ten (10) studies, which are the lifetime prevalence of domestic violence against women by their intimate partner. This correlates with the assailant's alcohol consumption, cigarette chewing, family history of violence, occupation, religion, educational status, residence, and decision-making power [50].

The National Demographic Health Survey Findings on Domestic Violence

The National Demographic and Health Survey is the nationally representative household survey of the Philippines, which provides valuable data from topics concerning basic demographic and health indicators which introduces a wide range of monitoring and impact evaluation indicators [51].

Data derived from the Philippines show that from the ever-married women consisting of ages 15 to 49 years old who reported ever experiencing physical, sexual, and/or emotional violence has declined over time with percentages ranging from 29% (gathered in 2008) to 18% (gathered in 2022)[52]. The decrease in numbers shows an improvement in reducing DV, considering that several interventions, such as laws on violence against women (VAW,) are implemented to address DV. However, these numbers do not equate to the diminishing of DV, as it is still both prevalent, inevitable, and a pervasive social issue in the country [46, 4]. Violence against women is considered by the United Nations as a *"human rights violation",* causing devastating effects such as long-term physical, sexual, and even mental consequences for women [3].

DV is recognized as a *"complex, multi-factorial phenomenon with dimensions that can include both intrinsic and extrinsic factors"* of which such determinants are identified to be an array of individual, biological, psychological characteristics, patriarchal ideologies, culture, society, religion, and media factors [48]. Articles from Lowe et al. and WHO support age as a factor DV, as these studies have consistently found that the early 20s to mid-30s is the age group wherein DV is commonly experienced by women [53, 46]. Age, being a demographic risk factor, suggests that at least 5 million acts of DV occur annually to women aged 18 years and older [45]. Despite the consistent occurrences of DV that are supported by these studies across the globe in the mentioned relevant age groups, Filipino women experience DV later in comparison with other countries [53].

Education is considered an intrinsic factor of DV. Supporting studies from Yoon & Bor, Puno et. al., Lowe et al., and Ahinkorah et al. show that women with low educational attainment have the highest risk of experiencing DV as they are economically less productive, seen to have less bargaining power within the household, and/or even subjugates of their husbands– putting these women in an inferior position to the opposite gender [53-57]. This is in comparison with established studies about DV and educational attainment of women, indicating that women possessing secondary or higher levels of education are significantly less likely to experience DV [58].

Bringing religion as a socio-cultural factor– this plays an instrumental role in DV, especially since Catholicism is embedded in the roots of the Filipinos since colonial times [59]. Studies support a significant relation between Catholicism and DV [55-56]. As biblical scriptures continue to teach women to be obedient, passive, and submissive to their husbands– this produces and contributes further to different forms of violence and–with the combination of the emergence of various religions and beliefs–their corresponding discourses, sacred texts, and own interpretations enacts exclusion and sexist discrimination of churches against women; gaining sustenance to fuel violence [60].

Vulnerability and continuous exposure to parental violence during one's formative years are also considered to have a long-lasting effect. Children who are exposed at least once in their lives are at greater risk for repeating the cycle of DV by either engaging in abusive relationships or becoming abusers [61]. Such circumstances are exhibited in a study with college students as respondents, indicating their reports of abuse as a contributor to their experiences with domestic violence in their current dating relationships [62]. In addition to this, children who were exposed to acts of DV are considered the second highest predictor for undergoing DV as an adult [63]. Agustin et al. found that children exposed to any form of violence have significant risks of undergoing alcohol or drug abuse, post-traumatic stress disorders, and/or even juvenile delinquency and/or adult criminality. DV has long-term implications for children, regardless of whether they are the ones experiencing DV or not, as these children suffer emotional and psychological trauma, as well as being denied the kind of household that fosters a healthy relationship and development [64].

An unfortunate factor would be the justification of women for DV. Circumstances such as alcohol abuse leading to aggression [46, 65], social identity leading to hostility [66], and even the comfort of going back to their abusers despite the present of logical reasoning that DV is wrong truly imparts to the violence inflicted among women by men [67]. Longevity of the relationship also plays a part, as women find it progressively difficult to leave, especially upon experiencing abuse within the first year [67]. Fortunately, Filipino women condemn any form of justification for DV. According to Clamor, influencing variables such as educational attainment, exposure to various kinds of media, and awareness of laws and help desks regarding DV are found to be significantly correlated to condemning domestic violence, lowering odds for Filipino women to discern domestic violence as acceptable and/or justifiable [68]. This supports the study of Salting, affirming that the awareness of Filipino women of help desks and other media avenues lowers the experience of DV [69]. Moreover, justification of DV can also be due to numerous existing social identities. In the Philippines, DV was found to be a repercussion of intergenerational transmission [55]. For instance, a child exposed to and witnessing DV with parents not only results in psychological and emotional trauma, but these children may also develop internalized gender expectations,

which unfortunately, may continue into adulthood [48]. This consequently leads to a belief that men are justified in having violent tendencies towards women, as these children recognize DV in parts of their childhood.

On the other hand, environmental factors and current life circumstances contribute to DV such as place of residence (either urban or rural) and household size. The Philippines shows lower prevalence in rural areas, in comparison to countries like Bangladesh and Pakistan [70]. In the context of household size, the presence and the continuous increase of living children increases the involvement of spousal violence [69, 71].

Women's empowerment has been found to have a direct impact on domestic violence in various contexts, including work status, wealth status, and decision-making [68, 69]. The lived experience of economic stress, living in a low-income country, being categorized in the lowest wealth quantile, lack of or limited access to necessities, falling short of work expectations, money, and status problems further increases the exposure of women to domestic violence.

Predicting the Significant Factors and Vulnerability of Women to Domestic Violence Using Machine Learning Algorithms

A comparative study conducted in Liberia utilized machine learning algorithms such as Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Random Forest (RF), Decision Trees (DT), XGBoost, LightGBM, and CatBoost to predict domestic violence vulnerability among Liberian women [72]. In this study, findings suggest that LightGBM and Random Forest models achieved the highest accuracy rates of 81% and 82%, based on the 2019-2020 Liberian Demographic Health Survey, respectively. Among the analyzed results, the number of women who had experienced emotional violence emerged as a significant predictor across multiple machine learning algorithms.

Basaran & Duru also performed a related study which discusses the frequency and risk factors of domestic violence against women using ML methods/algorithms. This study employed 6 machine learning algorithms, namely: decision tree, random forest, support vector machine (SVM), logistic regression (LR), Naive Bayes, and KNN [73]. Using these ML methods, the results show that from December 2023 to February 2024, over 11% of the 644 Turkish women have experienced domestic violence. The identified significant risk factors are duration of marriage, number of children, and level of education, while threat, insult, and injury were the common risk factors in the 6 machine learning algorithms. With a 100% accuracy rate, support vector machine (SVM) and logistic regression (LR) are seen as the models with the highest accuracy in predicting domestic violence against women in Turkey.

IPV or Domestic Violence (DV) is one of the crucial factors that contributes to South Africa's problem with violence that commonly leads to femicide, where killing a woman or girl because of her gender takes place [74-75]. As a result of this, Amusa et al. (2020) performed a study by predicting the vulnerability of South African women to IPV through tree-based machine learning techniques. Using the dataset of the 2016 South African Demographic and Health Survey, the author utilized four ML models, such as RF, LR, DT, and GB. The results show that DT and RF models dominated among the four ML models. However, the DT model was desirable for prediction since it has higher accuracy and higher sensitivity.

Moreover, a study on the prediction of Domestic Violence in Bangladesh during the onset of COVID-19 of which the researchers integrated the use of an oversampling technique called Synthetic Minority Sampling Technique (SMOTE) and chi-square statistical test to address class imbalances and utilized Random Forest, Logistic Regression, and Naive Bayes as their respective machine learning models along with relevant evaluation metrics such as Accuracy, Precision, Recall, F1-Score and further analysis using ROC-AUC and Confusion Matrix [76]. As a result, the researchers stated that family income level during the COVID-19 pandemic and education level of the family members are significant predictors of Domestic Violence in Bangladesh during COVID-19, with Random Forest being the best-performing model.

METHODOLOGY

Research Design

This study explores a quantitative, analytical, and comparative approach between the machine learning algorithms namely: K-Nearest Neighbors (KNN), Light Gradient Boosting Machine (LightGBM), Logistic Regression, Random Forest, and Xtreme Gradient Boosting (XGBoost) and their predictive abilities towards determining the vulnerability of Filipino Women to Domestic Violence, along with its significant risk factors.

Data Collection

The data for this study was obtained from the 2022 NDHS and was collected by the PSA, published in The DHS Program website of which the researchers derived authorization. This survey was conducted from May 2nd to June 22nd, 2022—the 12th series of the NDHS Survey. It includes data from approximately 27,821 Filipino women aged 15 to 49, categorized as Individual Recode (IR) data. Furthermore, the dataset comprises of 5,539 features, captures socio-demographic profiles, background characteristics, attitudes, current life circumstances, and immediate precursors of domestic violence in the Philippines.

Research Data

The data of this study explores two significant variables, the Independent and Dependent Variable. The Independent Variable of this study is based on the individual recode features of the dataset, varying from the socio-demographic factors, background characteristics, attitudes and viewpoints, as well as current life circumstances and immediate precursors of domestic violence in the Philippines. On the other hand, the study's dependent variable is split into two categories–Yes (value count: 1) and No (value count: 0)– and are separated in the 2022 NDHS through the individual recode variable "selected for domestic violence module" or variable V044, which is accounted for women who were or were not selected randomly by the DHS to answer for the domestic violence module for further assessment on reported cases/instances of domestic violence.

Data Pre-processing

To enhance data quality and ensure compatibility with the chosen algorithms, precise data pre-processing procedures were undertaken prior to the training of machine learning models. The researchers utilized Python Toolkits namely: Pandas, NumPy, Matplotlib, Seaborn, and Sci-Kit Learn which are core libraries that are considered specialized in Exploratory Data Analysis (EDA). Data cleaning was used to drop NaN values and K-Nearest Neighbors to impute missing values. Feature engineering was also deployed such as removal of constant features, Chi-Square Test of Independence for categorical variables, and dropping the highly correlated features through Spearman's Rank Correlation. Mutual Information was also utilized to further filter and rank the features. After employing several methods for feature engineering, the researchers opted to use domain knowledge to drop the remaining redundant and unnecessary features.

With this, the final remaining features total to 22 features, further categorized by the researchers for easier distinction, of which include the following:

1) **Household Demographics Features:** Age in 5-year groups, Number of household members (listed), Number of eligible women in the household (de facto), Sex of household head, Age of household head, Sons at home, Daughters at home, Owns a house alone or jointly
2) **Reproductive History and Awareness Features:** Births in last 5 years, Knowledge of ovulatory cycle, Index last child prior to maternity-health, Can women get pregnant after birth and before period
3) **Contraceptive Use and Needs Features:** Ever used anything or tried to delay or avoid getting pregnant and Unmet need for contraception

4) **Healthcare Access and History Features:** Visited health facility in the last 12 months and Had genital sore/ulcer in the last 12 months.
5) **Sexual Behavior and Activity Features:** Age at first sex, Recent sexual activity, Number of sex partners (including spouse) in the last 12 months, and Presence of children <10 during sexual activity section
6) **Technology Use for Media Awareness Feature:** Frequency of using internet last month
7) **Union Status Feature:** Current/formerly/never in union

Algorithm Testing and Hyperparameter Settings

The models were processed using a 10-Fold Stratified Cross-Validation and a hybrid sampling technique of SMOTE-Tomek to address class imbalances. The study's hyperparameter tuning was conducted through 3 cycles of Randomized Search Cross-Validation to identify optimal settings. These steps collectively ensured enhanced model performance, reduced overfitting, and provided fair, averaged comparisons.

Table 1. *Model's Best Parameters and Values*

Parameter	Values
K-Nearest Neighbors (KNN)	
Number of Neighbors	6
Weights	Distance
Distance Metric	Manhattan
Light Gradient Boosting Machine (LightGBM)	
Learning Rate	0.01
Subsample	0.6
Num_leaves	63
N_estimators	100
Min_child_samples	5
Max_depth	3
Colsample_bytree	1.0
Logistic Regression	
Penalty	Elasticnet
L1 Ratio	1.0
Inverse Regularization Strength	0.017670169402947963
Random Forest	
N_Estimators	500
Max_Depth	2
Min Samples Split	2
Criterion	Gini
Xtreme Gradient Boost (XGBoost)	
Learning Rate	0.1
Max_depth	5
Subsample	0.8
Scale_pos_weight	5
Colsample_bytree	0.6
Gamma	5
N_Estimators	100

Evaluation Metrics

Accuracy. Accuracy provided a quick classification figure, suitable for easy analysis, determining the model's performance in terms of predictive abilities.

$$\text{Accuracy} = \frac{\text{No. of correct predictions}}{\text{Total no. of predictions}}$$

Confusion Matrix. This evaluation metric is a performance measurement for machine learning classification problems wherein the output can be 2 or more classes. The confusion matrix of this study has 4 important terminologies, as stated below:

1) **True Positive (TP):** The model correctly predicted women who were vulnerable to domestic violence, accurately predicting that they were at risk.
2) **False Negative (FN):** The model incorrectly predicted that a woman was not vulnerable to domestic violence, even though she was actually at risk. The model fails to predict women who were, in fact, at risk.
3) **False Positive (FP):** The model incorrectly predicted that a woman was vulnerable to domestic violence when she was not at risk. The model mistakenly predicted women who were not at risk.
4) **True Negative (TN):** The model correctly predicted that a woman was not vulnerable to domestic violence, accurately predicting that they were not at risk.

F1 Score. The harmonic mean between recall and precision. This identified how precise and robust the classifier was.

$$F1 = 2 \star \frac{1}{\text{Precision} + \text{Recall}}$$

Precision. The measure of a model's performance, quantifying how many correct positive predictions the model produced.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall. An evaluation metric that measures how well a model identifies positive instances.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Log Loss. Evaluates the accuracy of probabilistic predictions, which penalizes wrong predictions with higher certainty more heavily.

$$Log\ Loss = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)$$

Where,

y_i is the true label and

p_i is the predicted probability of the positive class

Matthews Correlation Coefficient (MCC). A measure of the quality of binary classifications, considering all 4 confusion matrix categories.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

PRC-AUC (Precision-Recall Curve - Area Under Curve). Focus on the performance of the classifier with respect to the positive class.

$$PRC - AUC = \int_0^1 Precision(r)dr$$

ROC-AUC (Receiver Operating Characteristic - Area Under Curve). Evaluates a model's ability to distinguish between classes across all classification thresholds.

$$ROC - AUC = \int_0^1 TPR(x)dx$$

Ethical Considerations

Ethical considerations were highly acknowledged in this study. All data derived from the DHS Program was anonymized through case identifiers to protect privacy and its distribution is restricted. To mitigate model bias, variables were thoroughly reviewed based on existing literature, and fairness techniques were applied. We emphasize that these AI tools are intended to support, not replace, human decision-making in prevention. Lastly, this study identifies significant association of domestic violence risk factors through mathematical means, without establishing a cause-and-effect relationship.

RESULTS AND DISCUSSIONS

A. Model Performance Evaluation

Table 6.
Model Performance Comparison

Model	Evaluation Metrics							
	Accuracy	Precision	Recall	F1 Score	Log Loss	MCC	ROC-AUC	PRC-AUC
KNN	72.10	73.97	92.64	82.21	N/A	24.39	77.92	88.16
LightGBM	77.34	79.07	91.72	84.92	46.05	42.18	86.68	93.22
Logistic Regression	77.57	79.41	91.51	85.02	48.10	42.93	86.94	94.63
Random Forest	76.54	81.11	86.97	83.78	43.54	42.07	85.99	94.48
XGBoost	77.06	80.66	88.37	84.28	49.91	42.90	86.28	94.55

Among the five chosen machine learning models, Logistic Regression emerged to be the best model, achieving the highest F1 score, accuracy, MCC, ROC-AUC, and PRC-AUC. Its high precision and recall implied consistent identification of vulnerable women, while low Log Loss indicated reliable probability predictions, maintaining robust performance despite class imbalance.

LightGBM and XGBoost also performed strongly. LightGBM excelled in recall—which helps identify vulnerable women effectively, and maintained strong precision, earning the second-highest F1 score. It also featured well-calibrated probabilities and robust ROC-AUC/PRC-AUC. XGBoost, while less calibrated, showed excellent overall balanced metrics, including high precision and recall crucial for accurately identifying at-risk women with few false alarms. Its strong MCC, nearly matches Logistic Regression, which highlights its reliability across both classes despite imbalance.

Random Forest achieved the highest average precision and lowest Log Loss, which represents excellent control over false positives and highly reliable probability predictions. Despite a slightly lower recall compared to Logistic Regression and LightGBM, it maintained a strong overall performance, effectively balancing the detection of actual domestic violence cases with the avoidance of false alarms. Its robust MCC, good ROC-AUC, and excellent PRC-AUC further highlight RF's strong class discrimination and sensitivity in identifying true positive cases, which is crucial for predictive efforts in social sciences.

Lastly, the KNN model showed fair overall performance. Despite achieving the highest recall, its accuracy and precision indicated some false positives. The F1 score is also lower compared to other models. Its non-probabilistic nature led to an extremely high number of Log Loss and was excluded from evaluation to avoid any form of bias against the model. Its

lower MCC highlighted reduced robustness, which may be due to the curse of dimensionality and a bias towards the majority class, limiting its ability to learn complex, non-linear relationships crucial for domestic violence prediction. While its ROC-AUC and PRC-AUC were acceptable, they were lower than other models in this study.

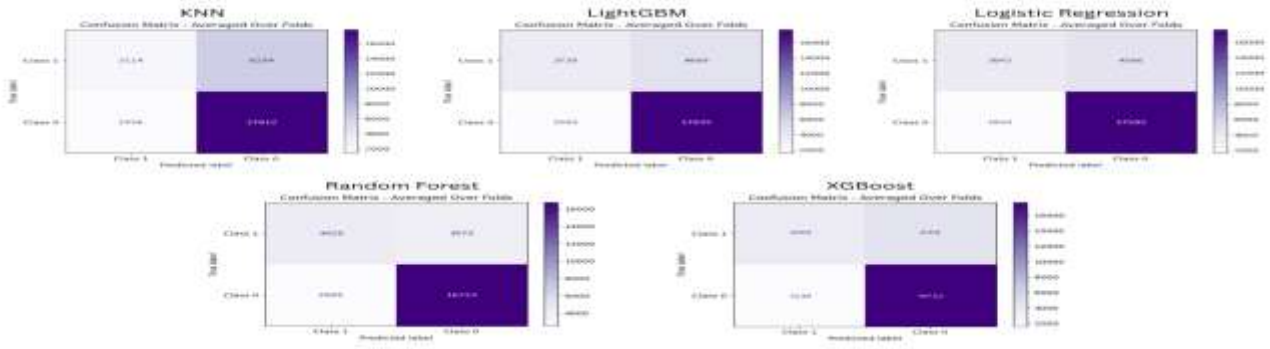


Figure 2. *Confusion Matrix of Machine Learning Models*

Figure 2 displays the confusion matrix comparison for the five machine learning models, visualizing their classification performance for women vulnerable to domestic violence. The goal for this section was to maximize True Positives (TP) while minimizing False Negatives (FN)

KNN showed a moderate performance, with a high number of FN (6,294), indicating a significant failure to identify vulnerable women.

LightGBM demonstrated strong performance (3,739 TP, 4,669 FN), with its FP (1,593) being a more acceptable trade-off than missed vulnerable cases.

Logistic Regression achieved balanced results, excelling in identifying non-vulnerable women (17,595 TN) and minimizing FP (1,633).

Random Forest produced high TP (4,429) and moderate FN (3,979), but its 2,505 FP may cause unnecessary concern.
XGBoost provided reliable results (4,305 TP compareD to 4,103 FN) and strong sensitivity, prioritizing the correct prediction of vulnerable women despite its 2,236 FP.

Feature Importance Analysis
Feature importance analysis was varied by model type. KNN used permutation-based methods which identified features whose randomization most disrupted performance. In tree-based models (LightGBM, Random Forest, XGBoost), importance was derived from the frequency and quality of splits involving each feature. Logistic Regression's feature importance was based on the magnitude and direction of its coefficients, indicating each variable's contribution to the model. Across these models, consistently high-importance features provide valuable insights into the predictive factors associated with the classification task and offer guidance for future data-driven recommendations.

European Journal of Computer Science and Information Technology, 14(8),13-35, 2026
Print ISSN: 2054-0957 (Print)
Online ISSN: 2054-0965 (Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

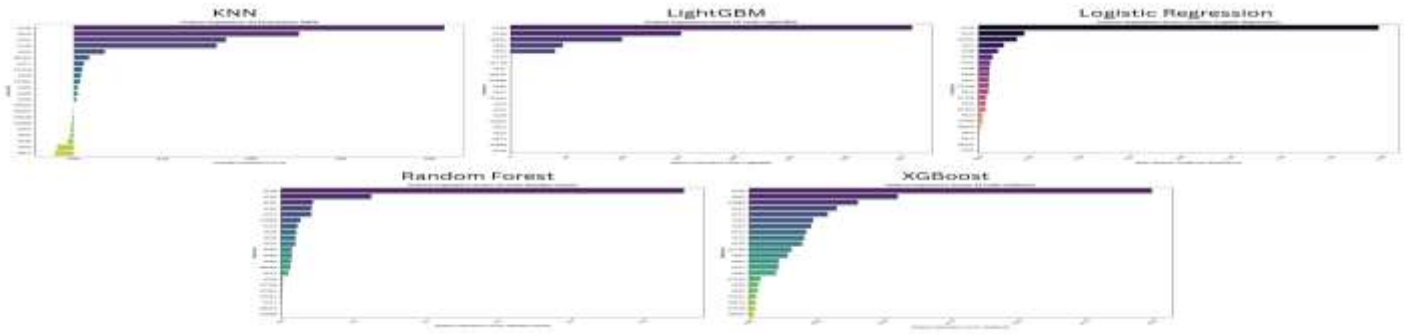


Figure 3. *Feature Importance Machine Learning Models*

Figure 3 represents the feature selection results which varied across all 5 models—utilizing the 22 derived features. LightGBM derived a concise set of 5 features. KNN derived 13 features. Logistic Regression and Random Forest both retained 20 features, and XGBoost incorporated all 22.

In-depth Correlation Analysis Between the Significance of the Dependent and Independent Variables
In-depth correlation analysis was conducted by the researchers to establish the specific associations between the derived features from the feature selection and the dependent variable. This analysis aimed to observe significant patterns and trends among features. Due to the binary nature of the dependent variable, the mean() function was used to represent the proportion of women experiencing domestic violence (class 1.0).

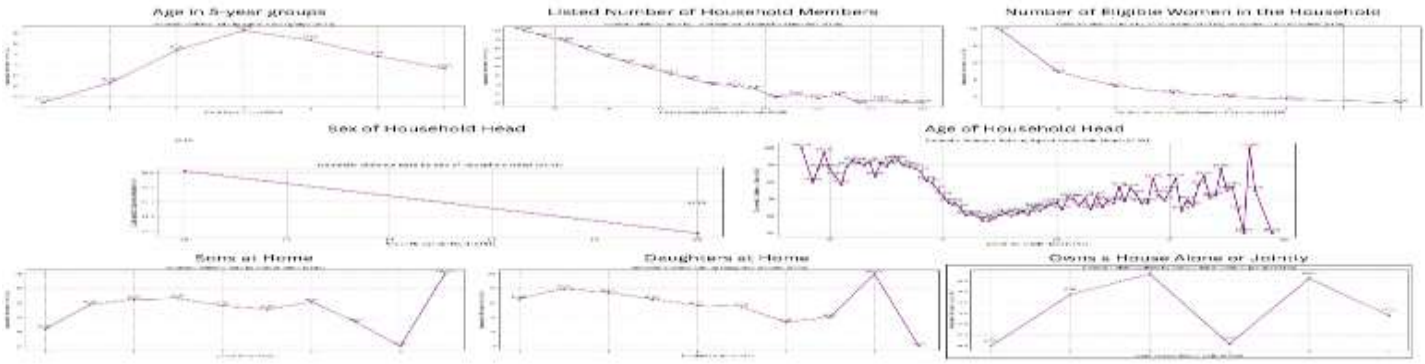


Figure 4. *Line Graph of Household Demographic Features*

Observing the figure above, the line graph primarily illustrates the relationship between the reported rate of domestic violence and the household demographic features.

Age in 5-year groups. The reported rate of domestic violence varied significantly with age. It was lowest for teenagers (15-19 years old) at 52.0%. Rates steadily increased through young adulthood, peaking at 86.2% for women in their early thirties (30-34 years old). A gradual decline was then observed for women aged 35 to 49, reaching 68.2% in the older age groups.

Listed Number of Household Members. This feature showed a strong inverse relationship with the reported rate of domestic violence. As the number of listed household members increased, the rate of domestic violence generally decreased. Notably, households with only one listed member exhibited a significantly higher rate of 100%.

Number of Eligible Women in the Household. This feature demonstrated a sharp inverse relationship with the reported domestic violence rate. Households with only one eligible woman showed a significantly high reported rate of 98.1%, with rates generally declining as the number of eligible women increased.

Sex of Household Head. This suggests a marginal difference, which denotes that when the household head is male, the reported rate of domestic violence is 70.1%, in contrast for female household heads of 67.9%.

Age of Household Heads. The highest rate is with the youngest heads of 15 years old (100%), remaining at an elevated pattern of 80% throughout the late teens and early 20s. Moreover, rates further decline on ages twenties to forties– reaching at its lowest rate in the late thirties to early 40s. Older households reported a fluctuating inconsistent pattern, between 60% to 80%. Lastly, a reported rate is further observed in households with very elderly heads– with an important definition of code 98—representing the unknown exact age of household head.

Sons at home. The reported domestic violence rate is 61.9% for sons not living at home. The increase in the number of sons suggests a gradual rise of 78.7% to 82.4%. This rating remains relatively high, with fluctuating trends of 82.4%, 83.1%, and 77.9%. Another increase is noted with 6 sons with a reported rate of 80.6%. The reported rate drops significantly to 66.7% with 7 sons in the household up to 8 sons with a rate of 50.0%. Lastly, the highest reported rate reached its peak of 100.0%, observed in households with 9 sons.

Daughters at home. This suggests a moderate domestic violence rate (65.6%) for households with no daughters with an increase rate of 79.5% with the presence of 1 daughter. A declining pattern was observed with 2 to 4 daughters in the household, ranging from 74.2%, 65.0%, and 56.7%, respectively. A stable rate was observed between 4 to 5 daughters and a drastic decrease with 6 daughters with a drop rating of 33.3%. 7 and 8 daughters are observed with a sharp increase in rates with 40.7% to 100% and another sharp downfall of 0.0% with 9 daughters.

Owns a house alone/jointly. House ownership indicated a 64.9% of reported domestic violence for women who do not own their house. A higher rate of 76.8%, present for women who owns their house alone. Its highest peak with 82.5% is observed among women who owns their house together with their partner or spouse. A lower rate is then observed (65.4%) for women who own their house jointly with someone else rather than their partner or spouse– exhibiting a similar percentage for those women who do not own any property. Meanwhile, for women who owns their house jointly with their partner and someone else, as well as owning solely with their partner, received a high rate of 80.5%. Lastly, a woman who's owning a house alone and jointly reported a moderate domestic violence of 72.0%.

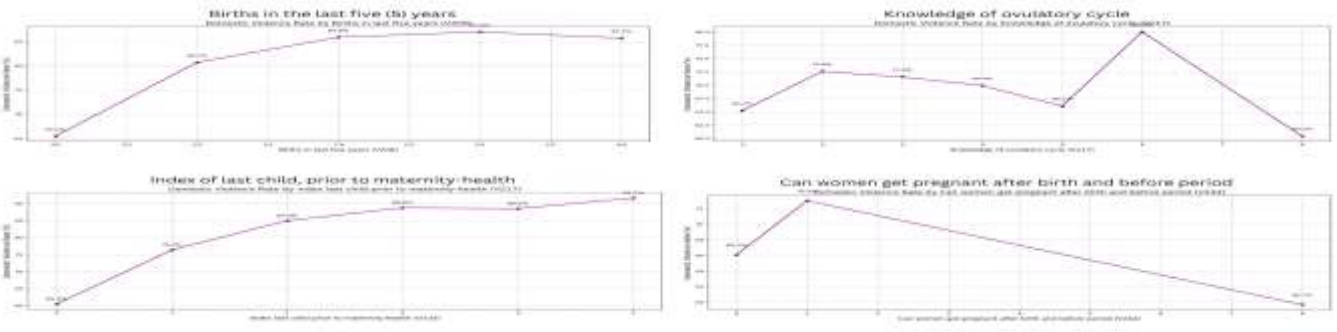


Figure 5. *Line Graph of Reproductive History and Awareness Features*

Observing the figure above, the line graph primarily illustrated the relationship between the reported rate of domestic violence and the reproductive and history awareness features.

Births in the last five (5) years. This suggests a rate of 65.5% for women who had no births in the past 5 years which increases to 80.7% for women who had 1 birth, and a further increase for women with 2 births. A high and stable rate is also observed for women who had 3 or 4 births in the last 5 years, with a reported rate of 87.0% and 85.7%, respectively.

Knowledge of the ovulatory cycle. This indicated lower rates for both women who did not know when they are most likely to get pregnant or believe that they can get pregnant anytime with a partnering percentage of 60.5% and 66.1% respectively. A higher rate is observed for women who believe that pregnancy is most likely to occur during their period with a domestic violence (65.2%). A moderate rate is suggested for women who think that they can get pregnant before their period (69.9%), forwarding to higher rates of 72.6% until 71.6% for women who believe fertility is highest after their period ends and that pregnancy can occur in the middle of their ovulatory cycle. Lastly, the highest reported rate is seen in the last category of women lacking knowledge on their ovulatory cycles.

Index the last child before maternity health. This indicates another inverse correlation between the rate of reported domestic violence with women having no births before the survey's reference period. Moreover, a gradual increase is observed, as the time since the last birth (up to 5 years prior) increases with the most significant reported rate occurring in the first year after the last birth. Lastly, the reported rate of domestic violence remains high and sustains stability for women whose last birth was 2 to 5 years– with a notable increase at 5 years prior.

If women can get pregnant after birth and before their period. This is observed to have a reported rate of domestic violence of 66.1% for women who believes in postpartum infertility. Categories of belief in the possibility of postpartum conception prior to the return of menstruation are reported to have the highest domestic violence rate of 73.0%. Lastly, women who are lacks certainty in postpartum fertility is observed to have the lowest reported rate of 59.7%.

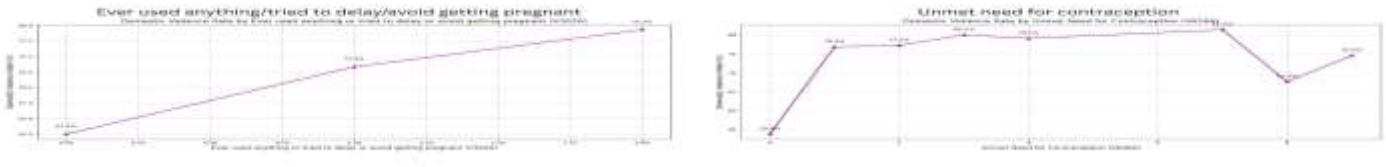


Figure 6. *Line Graph of Contraceptive Use and Needs Features*

Observing the figure above, the line graph primarily illustrates the relationship between the reported rate of domestic violence and contraceptive use and needs features.

Ever used anything/tried to delay/avoid getting pregnant. Women who reported neither current nor prior attempts to delay or avoid getting pregnant suggests the lowest domestic violence rate of 62.5%. Women who employed methods to avoid getting pregnant outside the survey's calendar period reported an increase in domestic violence, with a rating of 73.3%. Lastly, the highest peak was observed among women who had used or attempted to avoid getting pregnant, with a rating of 79.3%, within the survey's calendar period.

Unmet need for contraception. The relationship between unmet need for contraception and reported rate of domestic violence suggests that fertile women have the least reported rate of 53.8%. Higher rates are then observed among categories of women, particularly in line with during pregnancy, postpartum, and with the use of contraception. The highest reported rate of 80.1% is seen with women with no unmet contraceptive needs. Meanwhile, unmarried women and recently sexually inactive women have a lower rate of 67.7%, while a higher rate of 74.5% is found for women after the menopausal stage.

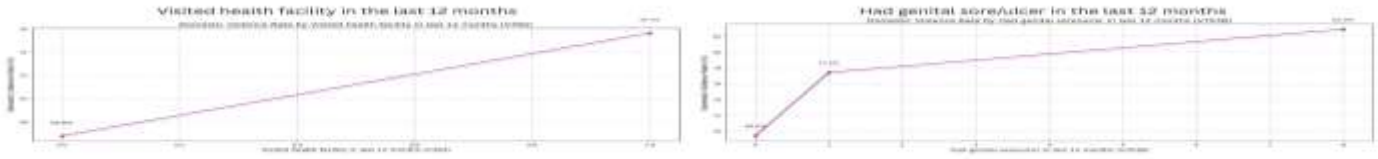


Figure 7. *Line Graph of Healthcare Access and History Features*

Observing the figure above, the line graph primarily illustrates the relationship between the reported rate of domestic violence and the healthcare access and history features.

Visited a health facility in the last 12 months. This suggests that women who did not visit a health facility in the past year, reported a lower rate of 66.8% compared to those who did with a rate of 75.7%.

Has genital sore/ulcer in the last 12 months. The highest reported rates are for women who had genital sore/ulcer in the last 12 months is under the category for women who lack knowledge regarding their status on this survey question, with a rating of 82.9%. Moreover, a reported rate of 69.4% is apparent for women who did not experience genital sore/ulcer in the last 12 months, for the past year, with an increased rate of 77.4% for women who had such a condition.

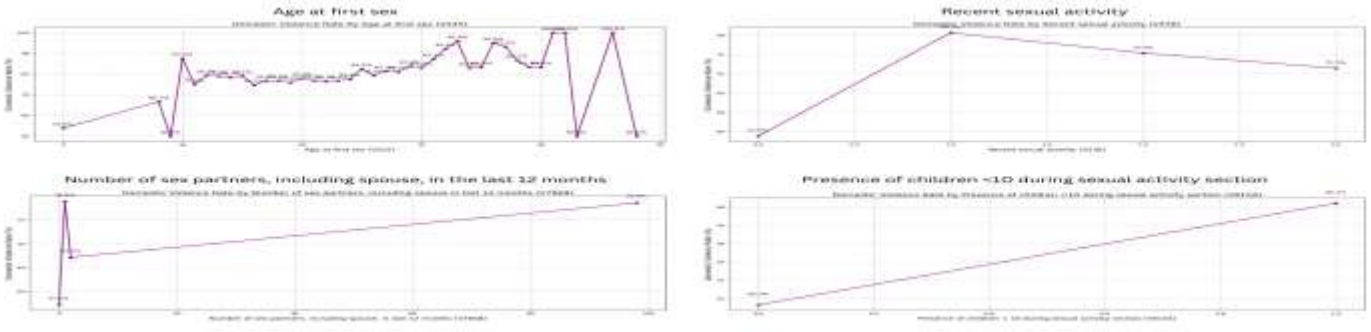


Figure 8. Line Graph of Sexual Behavior and History Features

Observing the figure above, the line graph primarily illustrates the relationship between the reported rate of domestic violence and the healthcare access and history features.

Age at first sex. Women who never had sexual intercourse reported a domestic violence rate of 53.8%. An increasing trend is noted for women whose first sexual experience occurred at younger ages (66.7% to 87.5%), around the age of 8 and age 10, followed by a fluctuating pattern in the late teens and early 20s. An elevated rate is seen in the late 20s to the early 40s with peak trends in mid-thirties. Lastly, inconsistent responses and lack of knowledge report a domestic violence rate of 100.0% and 50.0% respectively.

Recent sexual activity. The highest reported rate of domestic violence (80.7%) is observed among women who were sexually active in the 4 weeks, preceding the survey calendar. Women who reported never having sexual intercourse in between the survey calendar exhibited the lowest rate of 53.8%. For other categories such as women who were not recently sexually active (regardless of reason), abstaining due to postpartum reasons, and abstaining for other reasons reported an increase trend of 71.5% to 75.4%, For unmarried women who had not been sexually active for the past month and infecund, menopausal woman exhibited a domestic violence rate of 67.7% and 74.5%.

Number of sex partners, including spouse, in the last 12 months. The lowest domestic violence rate is observed among women with 0 sexual partners in the last 12 months. Women who reported having 1 sexual partner is represented to be the highest reported domestic violence rate. Women who had 2 sexual partners suggests a moderate rate of 67.2% while women who reported unknown number of sex partners reported a domestic violent rate of 78.6%.

Presence of children <10 during sexual activity section. The reported rate for women who reported no children being present during the sexual activity section depicts a rating of 69.3%. On the other hand, women for whom at least one child was present exhibited a higher rate of domestic violence of 80.4%.

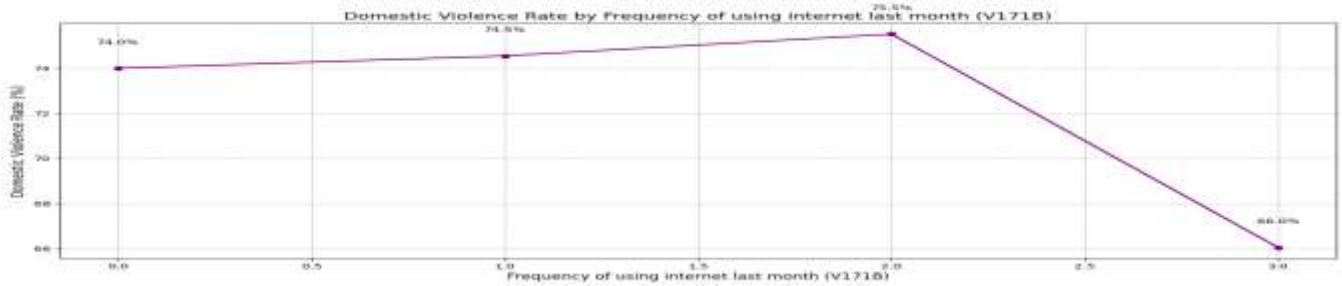


Figure 9. *Line Graph of Frequency of using internet last month feature*

Observing the figure above, the line graph primarily illustrates the relationship between the reported rate of domestic violence and the technology for media awareness feature.

Women who did not use the internet at all in the last month had a rate of 74.0. Women who used the internet less than once a week had a slightly higher rate at 74.5. For women who used the internet at least once a week, this category experienced the highest rate of domestic violence at 75.5%. Lastly, women who used the internet almost every day had a noticeably lower rate at 66.0%.

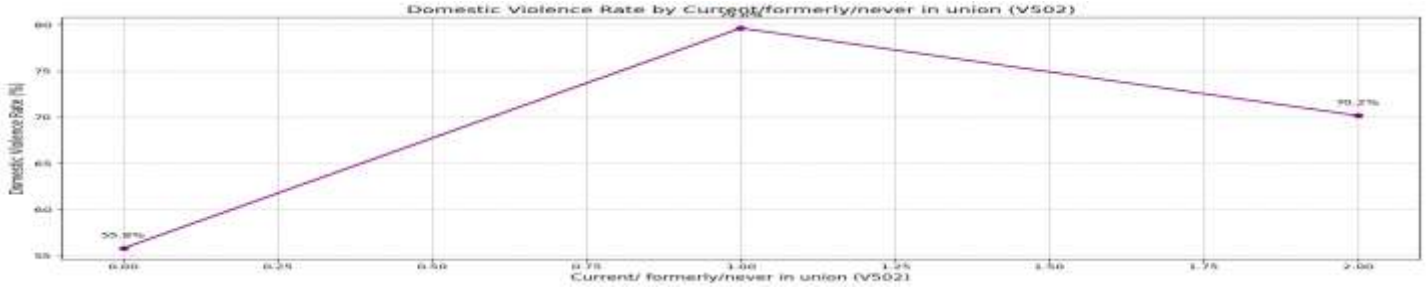


Figure 10. *Line Graph of Currently/Formerly/Never in union feature*

Observing the figure above, the line graph primarily illustrates the relationship between the reported rate of domestic violence and the union status feature.

Women who are currently in a union appear to have the highest reported rate (79.6%) compared to the other two groups. Women who have never been in a union suggest having the lowest reported rate (55.8%) compared to the other two groups. Furthermore, women who were formerly in a union appear to have a rate that falls in between the other two groups, with a rate of 70.2%.

SUMMARY

This study explored and compared the performance chosen machine learning algorithms in predicting the vulnerability of Filipino women to domestic violence through the utilization of a dataset derived from the 2022 NDHS. The researchers sought to determine which machine learning model could most accurately and reliably identify Filipino women at risk and identify significant factors related to DV. The methodology included a comprehensive pre-processing phase and processing phase of which the results were assessed using evaluation metrics in accordance with the algorithms used, to assess the performance of each model.

FINDINGS

Machine Learning Algorithms of the Study: K-Nearest Neighbors (KNN), Light Gradient Boosting Machine (LightGBM), Logistic Regression, Random Forest, and Xtreme Gradient Boosting (XGBoost) are chosen in this study to be the base machine learning algorithms based on the aspects of their interpretability and identification of risk factors, their robustness to dataset characteristics, as well as computational stability and efficiency that was partnered with various pre-processing techniques.

Model Performance: Logistic Regression emerged as the most reliable and appropriate model for predicting domestic violence vulnerability, significantly outperforming other algorithms. Its strength relies in its F1 Score (85.02%) and a high PRC-AUC (94.63%), which are both crucial metrics for imbalanced classification problems. This demonstrated the model's excellent balance between the recall and precision, ensuring reliable identification of at-risk individuals while effectively managing false positives. This robust performance is a meaningful contribution to data-driven approaches for addressing domestic violence in the Philippines, aligning with studies like Basaran & Duru (2024), which also utilized Logistic Regression effectively. While its high accuracy should be interpreted with caution due to class imbalance, the F1 Score and a strong MCC (42.93%) further validated the model's reliability across both classes. The ROC-AUC of 86.94% confirmed its strong overall discriminative power, while the outstanding PRC-AUC specifically highlighted its ability to maintain high precision across various recall levels, which is essential for accurately identifying true positive cases in a smaller, critical class.

Significant Factors: in terms of identifying the top significant key predictors of domestic violence vulnerability in Filipino women through feature importance analysis derived from the best performing model, Logistic Regression, which includes 22 features namely: Number of eligible women in the household (de facto), Index of last child prior to maternity-health, Presence of children <10 during sexual activity section, Daughters at home, Births in the last 5 years, Recent sexual activity, Sex of household head, Number of household members (listed), Visited health facility in the last 12 months, Currently/formerly/never in union, Had genital sore/ulcer in the last 12 months, Age in 5 year groups, Frequency of using internet last month, Sons at home, Owns a house alone or jointly, Age at first sex, Number of sex partners (including spouse) in the last 12 months, Ever used anything or tried to delay or avoid getting pregnant, If women can get pregnant after birth and before period, and Knowledge of ovulatory cycle, respectively.

In-Depth Analysis of Risk Factors: The analysis revealed several key predictors of domestic violence vulnerability in Filipino women through Feature Importance Analysis derived from the best-performing model, Logistic Regression, which includes the following:

1) **Household Composition:** Fewer household members were associated with significant risk, in contrast with the studies of Salting and Miedema & Kyaw [69, 71].
2) **Reproductive Knowledge and Autonomy:** Poor understanding of the ovulatory cycle and unmet need for contraception are significantly associated. This may potentially include underlying implications such as reduction of bodily autonomy, in line with decision-making abilities [68-69] and in contrast with Clamor's study regarding exposure to various media and awareness of reproductive laws/information [68].
3) **Socio-behavioral Factors:** Internet usage and joint home ownership patterns appeared as significant indicators, as seen from studies of Salting and Clamor supporting media exposure and law awareness and even lived experiences of women regarding their wealth quantile, and necessities [68-69].
4) **Demographic Attributes:** Age of the household head, particularly for young and elderly individuals, influenced risk levels. This may suggest potential economic instability that may arise to financial stress and an increase in violence, as supported by an additional study from Munir et al. [77] and even power imbalances or traditional gender roles which may suggest underlying support from studies of Sweet, Salting, and Clamor [68-69, 78]

CONCLUSION

In conclusion, this study employed five (5) machine learning algorithms, namely K-Nearest Neighbors (KNN), Light Gradient Boosting Machine (LightGBM), Logistic Regression, Random Forest, and Xtreme Gradient Boost (XGBoost) which is chosen by the researchers considering their interpretability and ability to identify risk factors, robustness to dataset characteristics, and computational stability and efficiency. Between them, Logistic Regression proved to be the best-performing model, especially in terms of the evaluation metrics F1 score, accuracy, MCC, ROC-AUC, and PRC-AUC. Moreover, among the 24 derived features from the extensive pre-processing methodologies, there are 22 remaining features which exhibit significant relationships based on feature importance analysis. And lastly, an in-depth analysis through line graph visualization between the dependent and independent variables (remaining features) were performed wherein the top key predictors of the Filipino Women's susceptibility to DV is further categorized to: Household Composition, Reproductive Knowledge and Autonomy, Socio-behavioral factors, and Demographic Attributes.

RECOMMENDATIONS

The recommendation of this study is further expounded and dedicated to two (2) sectors: policymakers and future researchers. For the concerned government agencies and policymakers, the researchers firstly recommend the following: expanding Violence Against Women (VAW) desk services up to the utilization of barangay-based surveillance; encouragement of family planning, maternal health programs, and even reproductive health awareness are probable options; inclusion of gender-based violence awareness and fertility misconceptions; age-group based interventions including economic assistance, awareness campaigns, economic assistance, and household management education; and the exploration of machine learning as a predictive tool.

To future researchers, it is recommended to improve data collection methods with safeguard, especially in accordance with the ethical considerations of this study. Additionally, data collection methods must be employed with safety and utmost precaution as samples are real-life individuals. The researchers also acknowledge the importance of Exploratory Data Analysis (EDA) and related studies for future researchers, as interpretable models should be used when deploying prediction tools in real-world settings. Although as a research gap, the study may lack studies locally– researchers may rely on international studies to further establish local support with proper methodology alongside. Another recommendation would

be the usage of other tests and methods such as a combination of performance metrics and statistical significance tests such as Paired T-Test for model comparisons or Friedman Test for evaluating multiple models across cross-validation folds, dependent upon the future study's statement of the problem. Lastly, if interpretability and causational analysis on the attributes of domestic violence are some of the objectives of future researchers, they may consider using SHAP values.

REFERENCES

1 Avdibegovic E, Brkic M, Sinanovic O. Emotional Profile of Women Victims of Domestic Violence. National Library of Medicine; 2017 Jun 29.
2 United Nations. What Is Domestic Abuse? [Internet]. Available from: https://www.un.org/en/coronavirus/what-is-domestic-abuse
3 UN Women. FAQs: Types of violence against women and girls [Internet]. UN Women Organization; 2024 Jun 27. Available from: https://www.unwomen.org/en/what-we-do/ending-violence-against-women/faqs/types-of-violence-against-women-and-girls
4 Philippine Commission on Women. Violence against women [Internet]. Available from: https://pcw.gov.ph/violence-against-women/#:~:text=According%20to%20the%202022%20National%20Demographic%20and,rape%2C%20and%201%2C505%20on%20acts%20of%20lasciviousness.
5 Lokanan M, Sharma S. WITHDRAWN: The use of machine learning algorithms to predict financial statement fraud [Internet]. ScienceDirect. Available from: https://www.sciencedirect.com/science/article/pii/S0890838924002051
6 Paula. Machine Learning Models for Precise Predictive Analytics [Internet]. Stefanini; 2023 Oct 4. Available from: https://stefanini.com/en/insights/news/machine-learning-models-for-precise-predictive-analytics
7 Chaw JK, Chaw SH, Quah CH, Sahrani S, Ang MC, Zhao Y, Ting TT. A predictive analytics model using machine learning algorithms to estimate the risk of shock development among dengue patients. Healthcare Analytics. 2024 Jun 1;5:100290.
8 Krichen M. Convolutional neural networks: A survey. Computers. 2023 Jul 28;12(8):151.
9 Wang B, Wu P, Chen Q, Ni S. Prediction and Analysis of Train Passenger Load Factor of High-Speed Railway Based on LightGBM Algorithm. Journal of Advanced Transportation. 2021;2021(1):9963394.
10 Craig L, Tucci L. " What is machine learning? Guide, definition and examples. s Interneta, https://www. techtarget. com/searchenterpriseai/definition/machine-learning-ML. 2024;15.
11 Siddique S, Haque MA, George R, Gupta KD, Gupta D, Faruk MJ. Survey on machine learning biases and mitigation techniques. Digital. 2023 Dec 20;4(1):1-68.
12 Fawcett EJ, Fairbrother N, Cox ML, White IR, Fawcett JM. The prevalence of anxiety disorders during pregnancy and the postpartum period: a multivariate Bayesian meta-analysis. The Journal of clinical psychiatry. 2019 Jul 23;80(4):1181.
13 Sarker IH. Machine learning: Algorithms, real-world applications and research directions. SN computer science. 2021 May;2(3):160.
14 Smelyakov K, Kitsenko Y, Chupryna A. Deepfake Detection Models Based on Machine Learning Technologies. In2024 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream) 2024 Apr 25 (pp. 1-6). IEEE.
15 Kulik J. How to implement predictive maintenance using machine learning? [Internet]. NeuroSYS; 2024 Apr 9. Available from: https://neurosys.com/blog/predictive-maintenance-using-machine-learning
16 Elhabyb K, Baina A, Bellafkih M, Deifalla AF. Machine learning algorithms for predicting energy consumption in educational buildings. International Journal of Energy Research. 2024;2024(1):6812425.
17 Mahmood A-B. Machine Learning Based Predicting Student Academic Success [Internet]. ResearchGate; 2020 Oct. Available from: https://www.researchgate.net/publication/344753535_Machine_Learning_Based_Predicting_Student_Academic_Success
18 Ahsan M, Khan A, Khan KR, Sinha BB, Sharma A. Advancements in medical diagnosis and treatment through machine learning: A review. Expert systems. 2024 Mar;41(3):e13499.
19 Ma Z, Mei G, Piccialli F. Machine learning for landslides prevention: a survey. Neural Computing and Applications. 2021 Sep;33(17):10881-907.
20 Rahman R, Khan MN, Sara SS, Rahman MA, Khan ZI. A comparative study of machine learning algorithms for predicting domestic violence vulnerability in Liberian women. BMC women's health. 2023 Oct 17;23(1):542.
21 Miesle P. What is the K-Nearest Neighbors (KNN) Algorithm? [Internet]. DataStax; 2025 Apr 7. Available from: https://www.datastax.com/guides/what-is-k-nearest-neighbors-knn-algorithm
22 Keylabs. K-Nearest Neighbors (KNN): Real-World Applications [Internet]. Keylabs: Latest News and Updates; 2024 Sep 13. Available from: https://keylabs.ai/blog/k-nearest-neighbors-knn-real-worldapplications/#:~:text=KNN's%20advantages%20include%20its%20simplicity,linear%20decision%20boundaries%2C%20and%20no
23 Raschka S. Why is Nearest Neighbor a Lazy Algorithm? [Internet]. Sebastian Raschka; Available from: https://sebastianraschka.com/faq/docs/lazy-knn.html
24 Ishwarya S. LightGBM: A Fast, Scalable, and Powerful Boosting Algorithm [Internet]. Medium; 2024 Sep 20 [cited 2025 May 25]. Available from: https://ishwaryasriraman.medium.com/lightgbm-a-fast-scalable-and-powerful-boosting-algorithm-497ffa4dce17

25 Khan AA, Chaudhari O, Chandra R. A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. Expert Systems with Applications. 2024 Jun 15;244:122778.
26 Yasar K, Lawton G, Burns E. logistic regression [Internet]. Search Business Analytics; 2024 Apr 25. Available from: https://www.techtarget.com/searchbusinessanalytics/definition/logistic-regression
27 Navlani A. Understanding Logistic Regression in Python [Internet]. DataCamp; 2024 Aug 11. Available from: https://www.datacamp.com/tutorial/understanding-logistic-regression-python
28 Govindarajan P, Abhishek M, Niharika K, Tesfahun A. Lungcanary-Pioneering Early Lung Cancer Detection using Machine Learning Algorithms. In2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS) 2024 Nov 7 (pp. 1-6). IEEE.
29 Awan AA. Naive Bayes classification tutorial using Scikit-Learn [Internet]. datacamp.com; 2023 Mar 3. Available from: https://www.datacamp.com/tutorial/naive-bayes-scikit-learn
30 Choudhary A. Naïve Bayes Machine Learning: Basics to Advanced Algorithm Concepts [Internet]. Medium; 2024 Nov 25. Available from: https://alok05.medium.com/na%C3%AFve-bayes-machine-learning-from-basics-to-advanced-algorithm-concepts-bcc4b585a6f9
31 Sruthi. Understanding random Forest Algorithm with examples [Internet]. Analytics Vidhya; 2025 Feb 28. Available from: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/
32 Donges N. Random Forest: A complete guide for machine learning [Internet]. Built In; 2024 Nov 26. Available from: https://builtin.com/data-science/random-forest-algorithm
33 IBM. What is Support Vector Machines (SVM)? [Internet]. [Place unknown]: IBM; 2023 Dec. Available from: https://www.ibm.com/think/topics/support-vector-machine
34 Anshul. Support Vector Machine (SVM) [Internet]. Analytics Vidhya; 2025 Apr 21. Available from: https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/
35 Kharkhar D. Unravelling the Power of XGBoost: Boosting Performance with Extreme Gradient Boosting [Internet]. Medium; 2023 Jul 30. Available from: https://medium.com/@dishantkharkar9/unravelling-the-power-of-xgboost-boosting-performance-with-extreme-gradient-boosting-302e1c00e555
36 Zhan C, Li J, Jiang W, Sha W, Guo Y. E-commerce sales forecast based on ensemble learning. In2020 IEEE International Symposium on Product Compliance Engineering-Asia (ISPCE-CN) 2020 Nov 6 (pp. 1-5). IEEE.
37 Alipio C. Domestic violence and migration in the Philippines. International Institute for Asian Studies Newsletter. 2014;67.
38 Rafal M. Indigenous Women in the Philippines and Its Combat to Injustice: A Channel for Women's Reinforcement Operation. FEU IPDER Research Journal [Internet]. 2015. ;5(1):53–60. Available from: https://repository.feu.edu.ph/bitstream/handle/123456789/2281/IPDER%20Vol.%205%20No.%201%20p.53.pdf?sequence=1
39 Saldua ADI. The Role of Women from Pre-Hispanic to Spanish era [Internet]. Tonks History Word Press; 2016 Feb 26. Available from: https://tonkshistory.wordpress.com/2016/02/22/the-role-of-women-from-pre-hispanic-to-spanish-era/
40 Caliguia GI. Rethinking Bakla: A Keyword in Philippine Conceptual, Sexual, and Social History.
41 Connections for Abused Women and Their Children. Stigma and Domestic Violence [Internet]. CAWC; 2023 Feb 27. Available from: https://www.cawc.org/news/stigma-and-domestic-violence/
42 Mendoza DJ. The women's movement: Policy issues, influence and constraints. InRoutledge Handbook of the Contemporary Philippines 2018 Feb 19 (pp. 418-426). Routledge.
43 Baker McKenzie Firm. Legal provisions [Internet]. Available from: https://resourcehub.bakermckenzie.com/en/resources/fighting-domestic-violence/asia/the-philippines/topics/1legal-provisions
44 United Nations Children's Fund. Domestic Violence Against Women and Girls [Internet]. Florence (Italy): UNICEF Innocenti Research Centre; 2000 Jun. Innocenti Digest; No. 6. Available from: https://anrows.intersearch.com.au/anrowsjspui/bitstream/1/14296/1/digest6e.pdf
45 Huecker SL, Shreffler SE, Bruening SK. Domestic Violence. In: StatPearls [Internet]. St. Petersburg (FL): StatPearls Publishing; 2024 Jul 24. Available from: https://www.ncbi.nlm.nih.gov/books/NBK499891/
46 World Health Organization. Violence against women [Internet]. 2024 Mar 25. Available from: https://www.who.int/news-room/fact-sheets/detail/violence-against-women
47 Metz C, Calmet J, Thevenot A. Women subjected to domestic violence: The impossibility of separation. Psychoanalytic Psychology. 2019 Jan;36(1):36.
48 Parcon C, Makani A. Intimate Partner Violence Experiences of Filipino Women: Evidence from the Philippines DHS [Internet]. Rockville (MD): ICF; 2024. DHS Working Papers No. 205. Available from: https://dhsprogram.com/pubs/pdf/WP205/WP205.pdf
49 Kalokhe A, Del Rio C, Dunkle K, Stephenson R, Metheny N, Paranjape A, Sahay S. Domestic violence against women in India: A systematic review of a decade of quantitative studies. Global public health. 2017 Apr 3;12(4):498-513.
50 Semahegn A, Mengistie B. Domestic violence against women and associated factors in Ethiopia; systematic review. Reproductive health. 2015 Dec;12:1-2.
51 Demographic and Health Survey (DHS). The DHS Program: Survey Types [Internet]. Available from: https://dhsprogram.com/methodology/survey-Types/dHs.cfm

52 Philippine Statistics Authority. 2022 Philippine National Demographic and Health Survey (NDHS) [Internet]. Quezon City (Philippines): Philippine Statistics Authority; 2023 Jun. Available from: https://psa.gov.ph/system/files/main-publication/2022%2520NDHS%2520Final%2520Report.pdf
53 Lowe H, Mannell J, Faumuina T, Sinclair L, Tamanikaiyaroi L, Brown L. Violence in childhood and community contexts: a multi-level model of factors associated with women's intimate partner violence experience in Samoa. The Lancet Regional Health–Western Pacific. 2024 Jan 1;42.
54 Yoon GH, Bor J. Effect of education on the risk of gender-based violence in the Philippines. Philippine Journal of Health Research and Development. 2020 Sep 28;24(3):1-8.
55 Puno A, Kim R, Jeong J, Kim J, Kim R. Intergenerational transmission of intimate partner violence among women: Evidence from the 2017 Philippines National Demographic and Health Survey [Internet]. SSM Popul Health. 2023 Sep. Available from: https://doi.org/10.1016/j.ssmph.2023.101392
56 Ahinkorah BO, Dickson KS, Seidu AA. Women decision-making capacity and intimate partner violence among women in sub-Saharan Africa. Archives of Public Health. 2018 Dec;76:1-0.
57 Marium S. Women's level of education and its effect on domestic violence in rural Bangladesh. IOSR Journal of Humanities and Social Science. 2014;19(5):40-5.
58 Fageeh WM. Factors associated with domestic violence: a cross-sectional survey among women in Jeddah, Saudi Arabia. BMJ open. 2014 Feb 1;4(2):e004242.
59 Catholicism in the Philippines | Religion and Public Life at Harvard Divinity School. (n.d.). Religion and Public Life. https://rpl.hds.harvard.edu/faq/catholicism-philippines
60 Krob DB, Steffen L. Religious influence on education and culture: violence against women as common sense. Procedia-Social and Behavioral Sciences. 2015 Feb 12;174:2374-9.
61 Forke CM, Catallozzi M, Localio AR, Grisso JA, Wiebe DJ, Fein JA. Intergenerational effects of witnessing domestic violence: Health of the witnesses and their children. Preventive medicine reports. 2019 Sep 1;15:100942.
62 Rodriguez A, Venzor C. The Impacts of Exposure to Domestic Violence in Childhood That Leads to Experiencing Violence in Future Intimate Partner Relationships [master's thesis]. San Bernardino (CA): California State University, San Bernardino; 2022. from: https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=2495&context=etd
63 Kimber M, Adham S, Gill S, McTavish J, MacMillan HL. The association between child exposure to intimate partner violence (IPV) and perpetration of IPV in adulthood—A systematic review. Child abuse & neglect. 2018 Feb 1;76:273-86.
64 Agustin CP, Guadamor ML, Aquino IP. Domestic violence on children and its effects to their Socio-Educational behaviour. International Journal of Advanced Research in Management and Social Sciences. 2017;6(5):149-61.
65 Wilkinson DL, Hamerschlag SJ. Situational determinants in intimate partner violence. Aggression and Violent Behavior. 2005 Mar 1;10(3):333-61.
66 Flynn A, Graham K. "Why did it happen?" A review and conceptual framework for research on perpetrators' and victims' explanations for intimate partner violence. Aggression and Violent behavior. 2010 May 1;15(3):239-51.
67 Thacker S. What makes women justify domestic violence? [Internet]. THE ESTD; 2021 Dec 03. Available from: https://www.theestablished.com/community/identity/what-makes-women-justify-domestic-violence
68 Clamor WL. Justification of intimate partner violence in the Philippines: Attitudes towards wife beating among Filipina women. Asia-Pacific Social Science Review. 2018;18(1):14.
69 Salting PA. Women Empowerment and its Relationship to Spousal Violence: Evidences from the 2017 Philippines National Demographic and Health Survey. Bangkok, Thailand. 2020;1.
70 Stojetz W, Brück T. Exposure to collective gender-based violence causes intimate partner violence. Journal of Development Economics. 2023 Sep 1;164:103054.
71 Miedema SS, Kyaw AT. Women's intergenerational intimate partner violence and household child abuse in Burma (Myanmar). SSM-Population Health. 2022 Mar 1;17:101010.
72 Rahman R, Khan MN, Sara SS, Rahman MA, Khan ZI. A comparative study of machine learning algorithms for predicting domestic violence vulnerability in Liberian women. BMC women's health. 2023 Oct 17;23(1):542.
73 Başaran F, Duru P. Determining domestic violence against women using machine learning methods: The case of Türkiye. Journal of Evaluation in Clinical Practice. 2024.
74 Abrahams N, Mhlongo S, Dekel B, Chirwa E, Ketelo A, Lombard C, Mathews S, Labuschagne G, Martin L, Manganyi T, Gounden T. Decrease in femicide in South Africa: three national studies across 18 years. Cape Town: South African Medical Research Council. 2022.
75 European Institute for Gender Equality. Femicide [Internet]. Available from: https://eige.europa.eu/gender-based-violence/femicide?language_content_entity=en
76 Hossain MM, Asadullah M, Rahaman A, Miah MS, Hasan MZ, Paul T, Hossain MA. Prediction on domestic violence in bangladesh during the covid-19 outbreak using machine learning methods. Applied System Innovation. 2021 Oct 13;4(4):77.
77 Suleman D, Iqbal F. The Mediating Effect of Lack of Social Interaction and Financial Stress between COVID-19 and Domestic Violence: A Post-COVID Scenario of Pakistan.

78 Sweet PL. Clustered vulnerabilities: The unequal effects of COVID-19 on domestic violence. American Sociological Review. 2024 Jun;89(3):421-48.