Global Misogyny? Predicting Political Violence Targeting Women Across the Globe, 2018-2021

Boseong Yun

The University of Chicago

Abstract

The prevalence of non-sexual political violence targeting women has received little academic attention in political violence studies. The incorporation of non-sexual violence into political violence is important since existing literature conceals many ways in which political violence extends beyond sexual violence. In such contexts, this research leverages computational methods to determine whether political violence targeting women is systematically different from other political violence using logistic regression, LASSO, and random forest techniques. Using data from the Armed Conflict Location & Event Data Project (ACLED) between 2018 and 2021, this research finds that political violence targeting women is distinct from other political violence events across most continents. This finding has important implications that prompt further research on the potential mechanisms that drive political violence targeting women in a cross-national manner. This research prompts the following question: *What is it in the globe that produces homogenous violence patterns notwithstanding substantial cross-national differences in social, economic, and political contexts?*

1. Introduction

Gender has been an important subject of academic inquiry in political violence. Many studies have investigated gendered patterns and their implications in political violence across multiple contexts. Scholars have examined the relationship between gender and political violence in armed conflicts (Moser & Clark, 2001; Hudson, Caprioli, Ballif-Spanvill, McDermott, & Emmet, 2009), terrorism (Ortbals & Poloni-Staudinger, 2018; Harris & Milton, 2016), and politics (Hakansson, 2021) across many countries. These studies have revealed that gender becomes an important determinant of political violence that shapes its motives, types, and impacts across numerous settings. However, most previous literature investigating the relationship between gender and political violence has largely neglected the prevalence of nonsexual political violence targeting women. This fetishization of sexual violence in political violence has concealed ways in which gender becomes an important determinant of political violence beyond sexual violence (Meger, 2016). This conceptual departure from sexualized political violence into incorporating non-sexualized political violence into analysis prompts re-examination of fundamental mechanisms that drive political violence targeting women.

Nevertheless, few studies have comprehensively examined political violence targeting women through incorporating both sexual and non-sexual types of events into analysis. The intrinsic issue to political violence studies in obtaining high-quality data have posed substantial obstacles to conducting reproducible research (Gleditsch, Metternich, & Rugbger, 2014). Additionally, difficulties in tracking the actors and victims of political violence and categorizing political violence targeting women have generated setbacks in conducting systematic analysis (Bardall, Bjarnegard, & Piscopo, 2020). The resultant paucity of academic research on political violence targeting women has left the following questions unanswered:

1. Do political violence targeting women happen in haphazard or systematic manners? How are these events different from other political violence events?
2. Are political violence events targeting women are systematic across the globe? What drives the unusual similarities in political violence targeting women across the globe that transcends cross-national differences in social, economic, and political contexts.

Question (1) is important since it remains unclear whether political violence targeting women happens in systematic or haphazard manners. For instance, it is unclear whether perpetrators of political violence attack women *because they are women* (Bardall, Bjarnegard, & Piscopo, 2020). The incorporation of non-sexual violence into political violence targeting women makes it less clear whether there are overarching political incentives to target women. It is thus important to quantify whether there are systematic patterns in political violence targeting women. Question (2) is important since few studies investigated and found cross-national similarities in political violence targeting women using quantitative methods (Van de Vilert, Schwartz, Huismans, Hofstede, & Daan, 1999; Yodanis, 2004). This question is particularly important since political violence targeting women happens due to political incentives that are shaped by respective political contexts. It remains puzzling why there are cross-national similarities when there are substantial differences in social, economic, and political contexts. It is thus important to extend cross-national studies further and examine whether there are systematic cross-national similarities in political violence targeting women.

In such contexts, this study aims to find whether political violence targeting women are systematically different from others across the globe. Using comprehensive dataset from the Armed Conflict Location & Event Data Project (ACLED), this study leverages computational methods to determine whether political violence targeting women are systematically different from other events using logistic regression, the LASSO, and random forest algorithms. With classification based on the decision boundary trained on the distinguishable patterns in political violence events, high performance of these predictive models is used to indicate systematic gendered patterns in political violence. The findings from this research will further be used to guide future inferential research on the mechanisms that drive political violence targeting women.

Finally, this research has important academic and practical implications. This comprehensive analytical focus distinguishes itself from previous literature that has heavier emphasis on sexual violence. The theoretical departure from sexualized violence into encompassing non-sexualized violence into analysis can reveal whether there are other dimensions in which political violence targeting women may persist and thus exacerbate gender inequality. Also, this research has practical implications. It has been extensively studied that women's political participation brings substantive changes to the political and economic spheres in many regions (Beckwith, 2010; Hessami & da Fonseca, 2020). With growing female political participation, it is crucial to have comprehensive understanding about how political violence targeting women jeopardizes their political participation and thus exacerbate gender inequality in multiple ways.

2. Literature Review

The fetishization of sexual violence in political violence conceals many ways in which political violence extends beyond sexual violence (Meger, 2016). The majority of political violence events targeting women are non-sexual, and the number of these events has been increasing over the past years (Kishi, Pavlik, & Matfess, 2019). This excessive focus on sexual violence decontextualizes and homogenizes myriads of political violence events targeting women and conceals different theoretical rationale for political violence targeting women. However, there have been active academic discourses on sexual violence to identify disparate theoretical bases for sexual violence.

Two primary camps have attempted to explain the mechanisms behind sexual violence during wartime: 1) opportunistic and 2) weapon of war. The primary argument of the opportunist camp is that sexual violence happens due to inherent masculine culture that encourage either individual combatants or a group of combatants to commit sexual in political violence events (Kirby, 2012). This mode of thought attributes the root cause of sexual violence during wartime to unbalanced gender relations that can be remedied through enforcing stricter policies, norms, and code of conduct (Wood, 2009; Cohen, 2013). The weapon of war camp maintains that sexual violence is mainly used as a war tactic. The primary tenet of the weapon of war camp is the instrumentality of wartime sexual violence in advancing specific political objectives and incentives (Kirby, 2012). This line of thought relegates sexual violence to one of many tools used to advance political incentives and places its root causes to political contexts (Leatherman, 2011).

While these studies important provide conceptual bases, they do not sufficiently explain why gender becomes a withstanding determinant of both sexual and non-sexual types of political violence across the globe. The opportunist camp does not fully explain the prevalence of non-sexual events that target women. For instance, the opportunist camp does not fully explain why some perpetrators wield other non-sexual forms of political violence against women. The weapon of war camp does not fully capture why political violence targeting women becomes a withstanding war tactic across the globe. For example, the weapon of war camp does not fully explain why violence targeting women is consistently used as a war tactic across the globe when there are substantial differences. These analytical shortcomings suggest that the emphasis on sexual violence can be misleading to capture the full dimensions of political violence targeting women. These frameworks in understanding sexual violence may be limited in uncovering underlying mechanisms and thus raises the necessity for a holistic analytical approach that incorporates both sexual and non-sexual political violence into its scope.

This research thus contributes to the emerging literature that identifies political violence targeting women in a comprehensive manner. This research attempts to find systematic gendered patterns in political violence across the globe. Additionally, this research seeks to find regional heterogeneities in political violence targeting women.

3. Data

This study uses political violence data from the Armed Conflict Locations & Event Data Project (ACLED). The ACLED is a disaggregated data collection, analysis, and crisis mapping project supported by governments and universities that provides comprehensive and consistent data on political violence across the globe.

The ACLED provides a comprehensive overview on political violence events through recording political violence events in an event-based manner. The ACLED provides detailed definition and categorization of political violence events through its own taxonomy that consists of 3 general categories, 6 broad event types, and 25 sub-event types. The types of political violence events included in the dataset range from non-violent events such as peace agreement to violent events such as bomb explosions for instance. Please see ACLED codebook for more information about how the ACLED defines and categorizes political violence events.

The ACLED uses four data sources and four sourcing strategies to provide high-quality data on political violence. The ACLED defines and categorizes political violence events through colleting all reported data from the following four primary data sources: (1) Traditional Media, (2) Reports, (3) Local Partner Data, (4) Verified Social Media. The ACLED utilizes a wide arrange of data sources to increase the coverage of their data on political violence. Additionally, the ACLED uses the following four primary strategies to enhance sourcing quality control: (1) Source Control, (2) Continued Identification of New Sources, (3) Corrections, (4) Anonymization. These sourcing strategies include cross-referencing and fact-checking to ensure accuracy and consistency in their data. Using these two broad data sourcing and sourcing quality strategies, the ACLED maintains a principled methodology to provide comprehensive and consistent dataset on political violence across the globe. For more information about their methodology, please see ACLED Methodology.

This study uses two datasets from the ACLED between January 2019 and April 2018: 1) All reported political violence events that happened across the globe, 2) All political violence events that are classified as political violence targeting women from January 2018 to April 2021. The ACLED defines a political violence event as political violence targeting women when the majority of the victims are females or when the event specifically targets women's events such as Women's Movement or Women of Zimbabwe Arise (WOZA). This study adopts the ACLED's definition and categorization of political violence and political violence targeting women for all types of events. The regions included in this analysis are Africa, Middle East, Latin America & the Caribbean, Central Asia & the Caucasus, East Asia, Southeast Asia, South Asia, and Europe. The United States has been excluded from the analysis due to lack of sufficient data.

This study merges these two ACLED datasets and creates the *woman* variable that is coded as 1 if it is classified as political violence targeting women and 0 otherwise based on multiple common indicators. The dependent variable is *woman* and the primary independent variable is *sub_event_type*. The *woman* variable indicates whether a political violence event has targeted woman or not. The *sub_event_type* variable contains information about the type of political violence events such as bomb explosion or peaceful demonstrations. The variable *fatality* shows the number of fatalities in an incidence. The *inter1* and *inter2* variables show the identity of the actors involved in an event such as civilians, militia, and government among many others. There are two variables, *inter1* and *inter2*, that indicate the actors involved in an event because most political violence events happen between two groups of actors. For one-sided political violence events, the *inter2* variable is coded as 0. Except for the value of 0 in the *inter2* variable, there is no inherent ordering in *inter1* and *inter2*. For example, there is no difference whether we observe militia in either *inter1* variable or *inter2* variable. This study uses this merged dataset for the entire analysis. The following variables are used in this analysis:

Table 1: The Description of Variables

Variable Name	Description
woman	The categorical variable suggesting women targeted political violence that has been coded as 1 if true and 0 otherwise.
year	The year in which an event happened
country	The country in which an event happened
sub_event_type	The sub-type of an event
fatality	The number of fatalities in an event
inter1	The identity of the actors involved in an event
inter2	The identity of the actors involved in an event

4. Methodology

This study first divides the balanced data into training data and testing data at the ratio of 80% and 20% respectively. Afterwards, this study performs minority oversampling technique

(SMOTE) to handle class imbalance in the training data. This study then performs 5-fold cross-validation on the training data to evaluate its cross-validated performance across (1) logistic model, (2) LASSO model, and (3) random forest model. This study tunes the parameters for the lasso model and random forest based on the Area Under the Curve – Receiving Operator Characteristic (AUC ROC) performance metric prior to comparing across different models. This study then chooses the best model based on AUC ROC metric and fits the selected model on the entire training data and evaluates its performance on the test data.

4.1. SMOTE and class imbalance

SMOTE is an oversampling technique that creates new synthetic samples based on the K-nearest neighbor technique (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). Specifically, the SMOTE algorithm oversamples the minority class observations by randomly interpolating and selecting the near minority class neighbor based on feature space. The SMOTE algorithm is a popular technique used to handle class imbalance issues in machine learning (Fernández, Garcia, Herrera, & Chawla, 2018). This research utilizes this SMOTE algorithm to address the class imbalance issue in the dataset.

The data used in this research suffers from class imbalance where most events are not classified as political violence targeting women. This can be problematic because most predictive machine learning algorithms are designed for balanced class distribution and thus can provide deceivable results (Sun, Wong, & Kame, 2009). For instance, class imbalance in the dataset are prone to providing false high accuracy rates through predicting the majority class for every observation due to insufficient information learned about the minority class. This can be inappropriate because it is important for the model to have a well-balanced sensitivity and specificity in classification. Thus, this study adopts the minority oversampling technique (SMOTE) to handle class imbalance issues by balancing the training data for models to sufficiently learn about the characteristics of both classes.

4.2. Logistic Regression

A logistic regression is a standard classification algorithm used to predict one or more categorical outcomes based on the independent variables (Menard, 2002). The logistic regression is used to estimate the probability of occurrence of an event using the Logit link function. The cost function that logistic regression minimizes is:

$$\boldsymbol{J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}log(h\theta\left(x^{(i)}\right)) + (1 - y^{(i)})log(1 - h\theta\left(x^{(i)}\right))\ \right]}$$

This study uses logistic regression to classify whether political violence events are targeting women or not.

4.3. The LASSO Regression

The LASSO is a penalized regression model that penalizes redundant regression coefficients to improve its power (Tibshirani, 1996). The LASSO regression places the L2-norm loss function on regression coefficients to shrink redundant coefficients towards zero using the λ penalty parameter. This shrinkage improves the power and stability of the model. The cost function that the LASSO regression model minimizes is:

$$\boldsymbol{\hat{\beta}_{\lambda}^{Lasso} = argmin\left\{\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j X_{ij}\right)^2 + \lambda\sum_{i=1}^{p}|\beta_j|\right\}}$$

This study uses the LASSO regression model to classify whether political violence events are targeting women or not.

4.4. Random Forest

Random Forest is an ensemble learning method that uses multiple machine learning algorithms to improve its performance. Random forest improves its performance by combining randomized decision tress and aggregating their predictions (Breiman, 2001). In this classification setting, random forest reduces Gini index at each branch to obtain the best branch split. This model minimizes the Gini index where $\hat{p}_{mk}$ is the proportions of observations in the m^{th} region in the k^{th} class.

$$\boldsymbol{Gini(K) = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk})}$$

This study uses random forest to classify whether political violent events are targeting women or not.

4.5. Cross-validation

The optimal parameters for the LASSO model and the random forest model are selected through cross-validation. In this process, the training data are randomly divided into 5-folds of equal size datasets where each subset is further divided into a training set and a testing set that comprises 80% and 20% of each fold respectively. Afterwards, each training subset is used to fit a selected model, and each test subset is used to produce predicted outcome at different levels of parameters. This research finds the optimal parameters for the LASSO model and the random forest model based on the highest AUC ROC metric. In the LASSO model, this research tunes the cost parameter λ through cross validation. In the random forest model, this research tunes *mtry*, which is the number of predictors that is randomly sampled at each split when creating the trees, and *min_n*, which is the minimum number of data points in a node that are required for the node to be split further. The maximum number of trees has been set to 1000 trees to allow for a variety of tree forms (Probst & Boulesteix, 2017). The selected parameters are the following:

Table 2: The Selected Parameters for Lasso and Random Forest

Region	**Cost Penalty (Lasso)**	**Mtry (Random Forest)**	**Min_n (Random Forest)**
Africa	0.00281	4	30
Central Asia & the Caucasus	0.00121	8	20
East Asia	0.000687	8	22
Europe	1e-05	8	20
Latin America & the Caribbean	0.00121	6	30
Middle East	0.000518	8	27
South Asia	0.0016	8	27

5. Analysis

In this section, this study presents the results in three ways: model-level accuracy rates, model-level classification accuracy, and model-level accuracy via receiver operating characteristic (ROC) curves and area under the curve (AUC) calculations.

5.1. Predictive Accuracy

In this sub-section, this study shows the mean accuracy prediction rates for each region across all three specified models in Table 2. The baseline average accuracy rates have been assumed to be 50% to determine whether political violence targeting women happens in a random manner or systematic manner. Although there are differences in the mean accuracy rates across the models, all three models consistently show high accuracy rates compared to the baseline of 50% accuracy rates. Specifically, the average accuracy rates are 68% for logistic regression, 67.6% for LASSO regression, and 73.9% for random forest. The tuned random forest model shows the best performance in the mean accuracy rates. There is no discernable difference between the logistic model and the LASSO model in the mean accuracy rates.

Additionally, there are regional heterogeneities in the mean accuracy for each region. The mean accuracy rates are high for Africa, Central Asia & the Caucasus, Europe, Middle East, and Southeast Asia across all three models. Based on the random forest model, the average accuracy rates are 82.6% for Africa, 91.5% for Central Asia & the Caucasus, 89.5% for Middle East, and 73.6 % for Southeast Asia. The mean accuracy rates are low for Latin America & the Caribbean and South Asia. Based on the random forest model, the average accuracy rates are 56.6% for Latin America & the Caribbean and 58.2% for South Asia. The result is less clear for East Asia with the Random Forest showing significant improvements over the logistic and lasso models. Overall, the results are robust to regional differences and alternative forms of model specification.

Table 2: The Average Accuracy Rate by Model

Region	**Logistic Regression**	**Lasso Regression**	**Random Forest**
Africa	0.743	0.742	0.826
Central Asia & the Caucasus	0.88	0.876	0.915
East Asia	0.552	0.553	0.709
Europe	0.664	0.661	0.685
Latin America & the Caribbean	0.55	0.548	0.566
Middle East	0.852	0.851	0.895
South Asia	0.556	0.556	0.582
Southeast Asia	0.65	0.628	0.736

Average	0.68	0.676	0.739

Notes 1: The average accuracy rates have been calculated by summing up the average accuracy for each region and dividing by the number of regions. Mathematically, $\sum_{r=1}^{R} \frac{1}{n} Accuracy_r$, *where the subscript* $r = region$.

5.2. Classification Accuracy

In this sub-section, this study presents the overall confusion matrix for each model. The confusion matrix provides detailed information about classification accuracy such as sensitivity and specificity among many other performance metrics. In this section, this research uses sensitivity and specificity to measure the classification accuracy of the models. Sensitivity, also known as the true positive rate, measures the proportion of positives that are correctly classified as positive. Specificity, also known as true negative rates, measures the proportion of negatives that are correctly classified as negative. These two metrics are important measurements of classification accuracy in an imbalanced dataset (Banerjee, Dehnbostel, & Preissner, 2018). The following confusion matrices show the results for all regions by each model. The overall results show that all three models show high sensitivity rates and moderately high precision rates. For instance, this study witnesses 0.817, 0.821, and 0.752 sensitivity rates and 0.688, 0.685, and 0.773 precision rates for logistic, LASSO, and random forest models respectively. If you would like to see all confusion matrices for each region by each model, please see Appendix.

Table 3-1: The Average Confusion Matrix for Logistic Model

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	377135 (67.09%)	2550 (0.45%)
Targeting women	170968 (30.41%)	11418 (2.03%)

Table 3-2: The Average Confusion Matrix for Lasso Model

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	375836 (66.86%)	2502 (0.44%)
Targeting women	172267 (30.64%)	11466 (2.03%)

Table 3-3: The Average Confusion Matrix for Random Forest Model

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	333673 (75.64%)	2360 (0.53%)
Targeting women	97899 (22.19%)	7168 (1.62%)

5.3. Checking Model Accuracy: ROC Curve and Area Under the Curve

In this sub-section, this study plots the receiving operator characteristic (ROC) curve and presents the total area under the curve (AUC) for each model. The ROC curve is created by plotting the true positive rate on the y-axis and the false positive rate on the x-axis. The ROC curve shows the performance of a model at each level of threshold. The AUC indicates the probability of correctly classifying a random pair of observations.

Figure 1-1. The Average ROC Curve For Logistic Regression

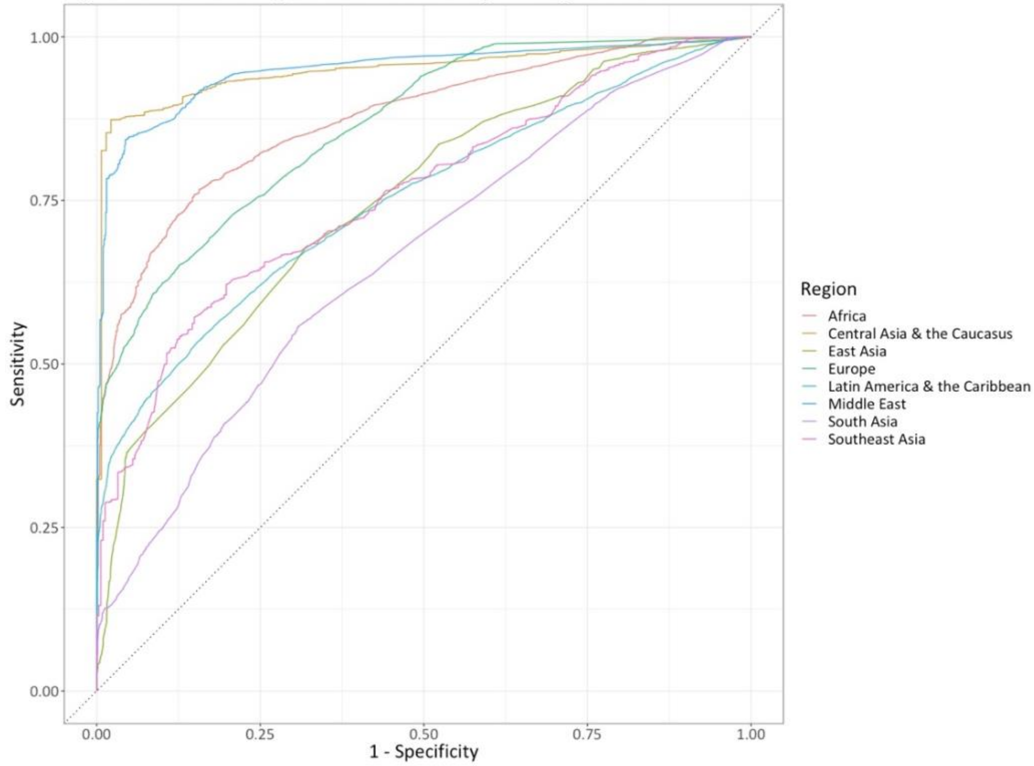
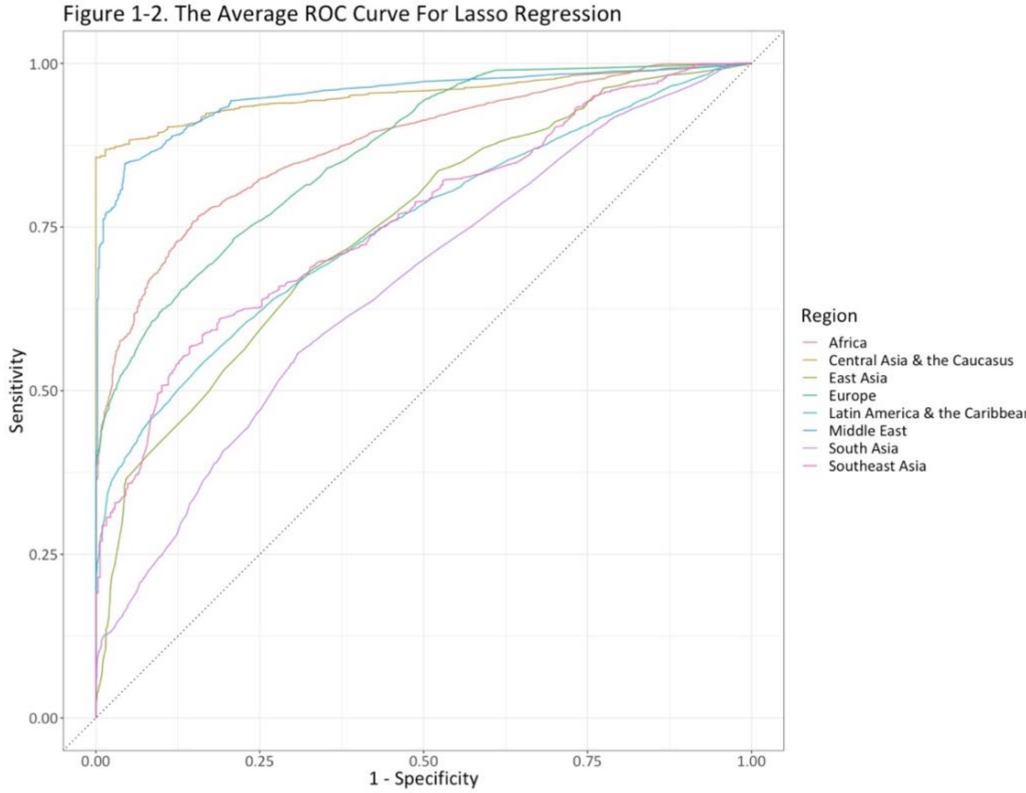
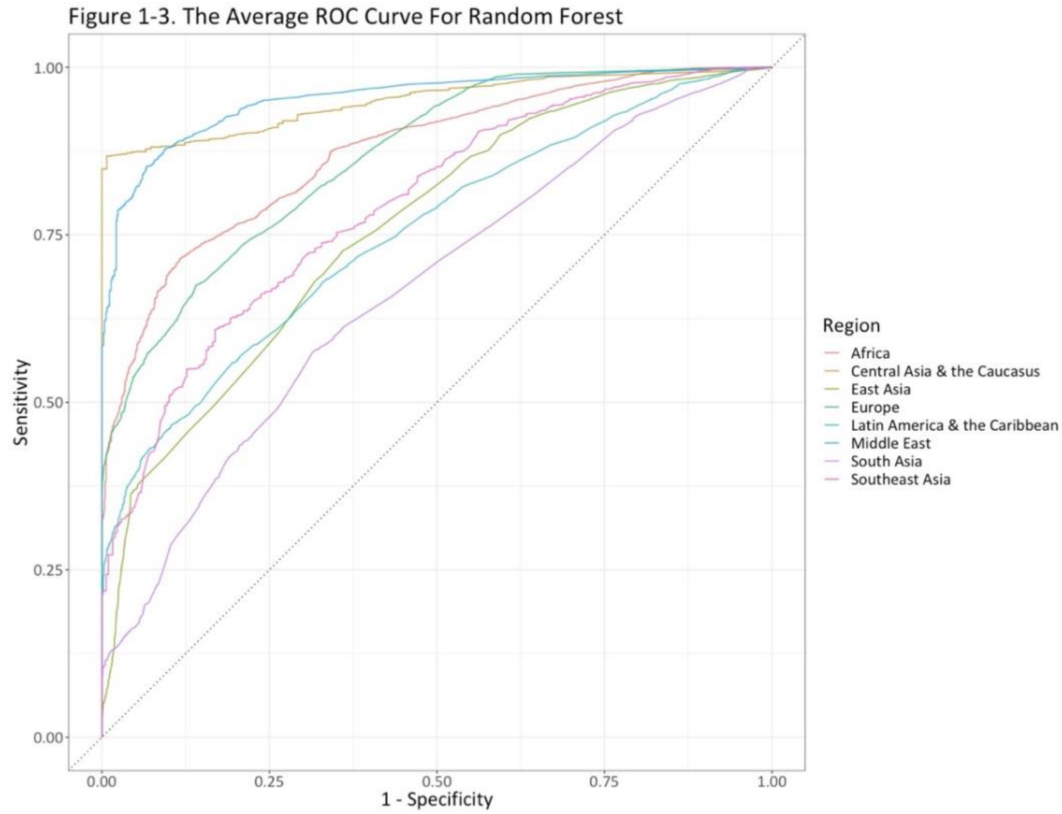


Figure 1-2. The Average ROC Curve For Lasso Regression





The following table shows the area under the curve in numerical from.

Table 4: The Area Under the Curve (AUC) by Model

Region	Logistic Regression	Lasso Regression	Random Forest
Africa	0.874	0.875	0.872
Central Asia & the Caucasus	0.95	0.954	0.951
East Asia	0.749	0.75	0.761
Europe	0.864	0.865	0.866
Latin America & the Caribbean	0.75	0.751	0.752
Middle East	0.95	0.952	0.954
South Asia	0.661	0.661	0.67
Southeast Asia	0.759	0.761	0.795
Average	0.819	0.821	0.827

These results show important results that there is strong consistency and agreement across all three models. This implies that the models have learned true gendered patterns in political violence.

5.4. Final fit

In this sub-section, this study selects the random forest model as the best performing model based on the AUC ROC metric and evaluates its performance on the test data. The results are almost identical to the results obtained using the cross-validation with the difference of 0.004 mean accuracy rates and 0.001 area under the curve.

Table 5: The Performance of the Final Fit from the Random Forest Model

Region	Accuracy	Roc_auc
Africa	0.812	0.86
Central Asia & the Caucasus	0.895	0.94
East Asia	0.725	0.762
Europe	0.68	0.869
Latin America & the Caribbean	0.566	0.757
Middle East	0.894	0.961
South Asia	0.582	0.661
Southeast Asia	0.73	0.817
Average	0.735	0.828

6. Discussion

The findings support that there are systematic gendered patters in political violence around the globe. Specifically, the predictive models used in this study showed high performance in classifying political violence events for most regions. Additionally, the results showed slight regional differences with lower performance for Latin America & the Caribbean and South Asia. In this section, this study analyzes the substantive implications of each result.

In the predictive accuracy section, this research has demonstrated that the best performing predictive model showed 74% accuracy rates in classifying whether political violence events are targeting women or not. This is a 24-percentage point increase from the baseline 50% accuracy rates. This finding indicates that there is a global pattern in political violence where political violence events targeting women have distinctive characteristics. Additionally, this research has shown regional differences where the model performed particularly worse for Latin America & the Caribbean and South Asia. This finding implies that these regions may not exhibit strong gendered patterns in political violence. This is an important discovery that can encourage more research on the existence of these regional heterogeneities.

In the classification section, this research provided additional information about the classification performance of the predictive models. This research showed that all three models showed 0.8 of sensitivity and 0.7 of specificity on average. This is an important finding because high and well-balanced sensitivity and specificity indicate that the model did not suffer from the false high accuracy rates problems addressed earlier for imbalanced datasets. Moreover, high performance in these two metrics showed that the models had small number of false negatives and false positives. As shown in the Appendix section 8-2 through 8-4 for confusion matrices, however, these results were less clear for Latin America & the Caribbean and South Asia where sensitivity was much higher than specificity. In the checking model accuracy section, the overall accuracy was further corroborated using ROC curves and AUC. The metrics showed strong coherence and consistency of the models in classifying political violence events. These results once again verify systematic gendered patterns in political violence across the globe.

7. Conclusion

In conclusion, this research has presented evidence of systematic gendered patterns in political violence around the globe. Although there were regional differences, the predictive models classified the political violence events correctly for most regions. Additionally, other measures such as sensitivity, specificity, and ROC AUC further corroborated evidence in favor of gendered patterns in political violence.

This research contributes to the emerging literature on political violence targeting women through finding systematic gendered patterns in political violence. This comprehensive analysis that incorporates both sexual and non-sexual political violence was important considering previous analytical limitations in their emphasis on sexual violence. Also, this research has found regional heterogeneities in classification accuracy rates and prompt further research. In short, this research has illuminated the necessity to establish political violence targeting women as a new strand of literature apart from sexual violence and political violence.

Additionally, this research has practical implications considering growing number of women participating in politics over the past decades. Studies have revealed that women's political participation bring substantive changes to the political and economic spheres. For instance, women's political participation changes public policies and public spending (Hessami & da Fonesca, 2020), shifts political institutions and representations (Beckwith, 2010), and encourages more female political participation (Liu & Banaszak, 2017). These results show that political violence targeting women that creates a high-risk political space for women can translate into gender inequality. Thus, this research highlights the existence of political violence targeting women across the globe and prompts further research and actions to prevent the violence. The findings of the research can also be used to channel more resources into combating non-sexual political violence events targeting women.

Finally, and most importantly, it is concerning to witness that there are global patterns in political violence targeting women. These patterns signify fundamental questions about the relationship between gender and political violence. What is it in the globe that produces homogenous violence patterns notwithstanding substantial cross-national differences in social, economic, and political contexts? What is in the human nature that has produced such clear demarcation in the patterns of political violence?

References

Banerjee, P., Dehnbostel, F. O., & Preissner, R. (2018). Prediction is a balancing Act: importance of sampling methods to balance sensitivity and specificity of predictive models based on imbalanced chemical data sets. *Frontiers in chemistry*, *6*, 362.

Bardall, G., Bjarnegård, E., & Piscopo, J. M. (2020). How is political violence gendered? Disentangling motives, forms, and impacts. *Political Studies*, *68*(4), 916-935.

Beckwith, K. (2010). A comparative politics of gender symposium introduction: Comparative politics and the logics of a comparative politics of gender. *Perspectives on Politics*, *8*(1), 159-168.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321-357.

Cohen, D. K. (2013). Explaining rape during civil war: Cross-national evidence (1980–2009). *American Political Science Review*, *107*(3), 461-477.

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research*, *61*, 863-905.

Gleditsch, K. S., Metternich, N. W., & Ruggeri, A. (2014). Data and progress in peace and conflict research. *Journal of Peace Research*, *51*(2), 301-314.

Håkansson, S. (2021). Do women pay a higher price for power? Gender bias in political violence in Sweden. *The Journal of Politics*, *83*(2), 515-531.

Harris, C., & Milton, D. J. (2016). Is standing for women a stand against terrorism? Exploring the connection between women's rights and terrorism. *Journal of Human Rights*, *15*(1), 60-78.

Hessami, Z., & da Fonseca, M. L. (2020). Female political representation and substantive effects on policies: A literature review. *European Journal of Political Economy*, *63*, 101896.

Hudson, V. M., Caprioli, M., Ballif-Spanvill, B., McDermott, R., & Emmett, C. F. (2009). The heart of the matter: The security of women and the security of states. *International Security*, *33*(3), 7-45.

Kirby, P. (2013). How is rape a weapon of war? Feminist International Relations, modes of critical explanation and the study of wartime sexual violence. *European Journal of International Relations*, *19*(4), 797-821.

Kishi, R., Pavlik, M., & Matfess, H. (2019). Terribly and Terrifyingly Normal': Political Violence Targeting Women. *Report for the Armed Conflict Location & Event Data Project*.

Leatherman, J. (2011). *Sexual violence and armed conflict*. Polity.

Liu, S. J. S., & Banaszak, L. A. (2017). Do Government Positions Held by Women Matter? A Cross-National Examination of Female Ministers' Impacts on Women's Political Participation. *Politics & Gender*, *13*(1), 132-162.

Meger, S. (2016). The fetishization of sexual violence in international security. *International Studies Quarterly*, *60*(1), 149-159.

Menard, S. (2002). *Applied logistic regression analysis* (Vol. 106). Sage.

Moser, C. O., & Clark, F. (2001). *Victims, perpetrators or actors?: Gender, armed conflict and political violence*. Palgrave Macmillan.

Ortbals, C. D., & Poloni-Staudinger, L. M. (2018). *Gender and Political Violence: Women Changing the Politics of Terrorism*. Springer.

Probst, P., & Boulesteix, A. L. (2017). To Tune or Not to Tune the Number of Trees in Random Forest. *J. Mach. Learn. Res.*, *18*(1), 6673-6690.

Sun, Y., Wong, A. K., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *International journal of pattern recognition and artificial intelligence*, *23*(04), 687-719.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288.

Van de Vliert, E., Schwartz, S. H., Huismans, S. E., Hofstede, G., & Daan, S. (1999). Temperature, cultural masculinity, and domestic political violence: A cross-national study. *Journal of Cross-Cultural Psychology*, *30*(3), 291-314.

Wood, E. J. (2009). Armed groups and sexual violence: When is wartime rape rare?. *Politics & Society*, *37*(1), 131-161.

Yodanis, C. L. (2004). Gender inequality, violence against women, and fear: A cross-national test of the feminist theory of violence against women. *Journal of interpersonal violence*, *19*(6), 655-675.

8. Appendix

8.1. Confusion Matrix for Each Region in Logistic Regression Model

Table 8.2-1: Confusion Matrix for Africa

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	48963	174
Targeting women	17160	1090

Table 8.2-2: Confusion Matrix for Central Asia & the Caucasus

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	46670	10
Targeting women	6356	127

Table 8.2-3: Confusion Matrix for East Asia

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	12914	408
Targeting women	11444	1681

Table 8.2-4: Confusion Matrix for Europe

Truth / Prediction	Not targeting women	Targeting women

Not targeting women	52888	372
Targeting women	27628	2349

Table 8.2-5: Confusion Matrix for Latin America & the Caribbean

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	58038	639
Targeting women	49613	3337

Table 8.2-6: Confusion Matrix for Middle East

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	103917	49
Targeting women	18079	744

Table 8.2-7: Confusion Matrix for South Asia

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	42711	820
Targeting women	34765	1860

Table 8.2-8: Confusion Matrix for Southeast Asia

Truth / Prediction	Not targeting women	Targeting women

Not targeting women	10934	78
Targeting women	5923	230

8.2. Confusion Matrix for Each Region in Lasso Regression Model

Table 8.3-1: Confusion Matrix for Africa

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	48895	173
Targeting women	17228	1091

Table 8.3-2: Confusion Matrix for Central Asia & the Caucasus

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	46535	6
Targeting women	6591	131

Table 8.3-3: Confusion Matrix for East Asia

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	12933	407
Targeting women	11425	1682

Table 8.3-4: Confusion Matrix for Europe

Truth / Prediction	Not targeting women	Targeting women

Not targeting women	52679	356
Targeting women	27837	2365

Table 8.3-5: Confusion Matrix for Latin America & the Caribbean

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	57803	628
Targeting women	49848	3348

Table 8.3-6: Confusion Matrix for Middle East

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	103733	41
Targeting women	18263	752

Table 8.3-7: Confusion Matrix for South Asia

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	42709	819
Targeting women	34767	1861

Table 8.3-8: Confusion Matrix for Southeast Asia

Truth / Prediction	Not targeting women	Targeting women

Not targeting women	10549	72
Targeting women	6308	236

8.3. Confusion Matrix for Each Region in Random Forest Model

Table 8.4-1: Confusion Matrix for Africa

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	54751	387
Targeting women	11372	877

Table 8.4-2: Confusion Matrix for Central Asia & the Caucasus

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	48656	36
Targeting women	4470	101

Table 8.4-3: Confusion Matrix for East Asia

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	13931	580
Targeting women	5571	1077

Table 8.4-4: Confusion Matrix for Europe

Truth / Prediction	Not targeting women	Targeting women

Not targeting women	43790	330
Targeting women	20622	1849

Table 8.4-5: Confusion Matrix for Latin America & the Caribbean

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	24001	306
Targeting women	19070	1275

Table 8.4-6: Confusion Matrix for Middle East

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	109220	102
Targeting women	12776	691

Table 8.4-7: Confusion Matrix for South Asia

Truth / Prediction	Not targeting women	Targeting women
Not targeting women	26895	518
Targeting women	19590	1091

Table 8.4-8: Confusion Matrix for Southeast Asia

Truth / Prediction	Not targeting women	Targeting women

Not targeting women	12429	101
Targeting women	4428	207