Forecasting Electoral Violence

David Randahl, Maxine Leis, Tim Gåsste,
Hanne Fjelde, Håvard Hegre, Staffan I. Lindberg,
and Steven Wilson

November 2024

Working Paper
SERIES 2024:150
THE VARIETIES OF DEMOCRACY INSTITUTE

UNIVERSITY OF GOTHENBURG
DEPT OF POLITICAL SCIENCE

Varieties of Democracy (V-Dem) is a unique approach to conceptualization and measurement of democracy. The headquarters – the V-Dem Institute – is based at the University of Gothenburg with 18 staff. The project includes a worldwide team with 5 Principal Investigators, 23 Project Managers, 26 Regional Managers, 134 Country Coordinators, Research Assistants, and 4,000 Country Experts. The V-Dem project is one of the largest ever social science research-oriented data collection programs.

Please address comments and/or queries to:

V-Dem Institute
Department of Political Science
University of Gothenburg
Sprängkullsgatan 19, Box 711
405 30 Gothenburg
Sweden
E-mail: contact@v-dem.net

V-Dem Working Papers are available in electronic format at https://www.v-dem.net

Forecasting Electoral Violence

David Randahl*[1], Maxine Leis[1], Tim Gåsste[1], Hanne Fjelde[1], Håvard Hegre[1,2], Staffan I. Lindberg[3], and Steven Wilson[4]

[1]Dep. of Peace and Conflict Research, Uppsala University
[2]Peace Research Institute Oslo
[3]V-Dem Institute, University of Gothenburg
[4]Dep. of Politics, Brandeis University

October 2024

*Corresponding author: David Randahl

Abstract

Electoral violence remains a significant challenge worldwide. It not only threatens to undermine the legitimacy and fairness of electoral outcomes, but often has serious repercussions on political stability more broadly. The ability to prevent electoral violence is critical for safeguarding democracy and ensuring peaceful transitions of political power. Predicting which elections are at risk of violence is an important step for effective prevention. In this study, we build and train a set of machine-learning models to forecast the likelihood of electoral violence on a global scale. Using a comprehensive set of data sources, with features including economic indicators, records of historical violence, political instability, and digital vulnerability, we predict the risk of electoral violence on a scale from no violence to severe violence. When combining a subset of these models to produce ensemble predictions of electoral violence for 2024-2025, our results show that our model effectively discriminates between the different levels of risk with a high degree of predictive accuracy. This research contributes to the field of political violence prediction by providing a medium-term data-driven forecasting tool for electoral violence. This knowledge may assist practitioners in the field of violence prevention by pinpointing elections at risk.

Keywords: Electoral violence Forecasting Elections Machine learning Political violence

1 Introduction

Today, a vast majority of countries in the world hold some form of election in which citizens vote to fill the highest political offices of the state. Where elections are free and fair, they represent a cornerstone of democratic governance that can provide a peaceful mechanism for transferring power and holding governments accountable.[1] Yet, electoral violence— i.e. violence seeking to purposefully influence the process or outcome of elections—represents a serious challenge to the integrity and legitimacy of electoral processes around the world (Birch, Daxecker, and Höglund, 2020). One in five national elections since 1946 has experienced significant levels of intimidation, harassment, and physical violence, often leading to civilian fatalities (Hyde and Marinov, 2012). In addition to the human suffering associated with all types of political violence, electoral violence risks undermining the legitimacy and fairness of the democratic process by influencing who stands for political office, who votes and whom they vote for, how votes are counted, and how electoral outcomes are enforced. The potentially far-reaching implications for political stability and social cohesion may even trigger broader political turmoil, including civil wars (Birch, Daxecker, and Höglund, 2020; Birch and Muchlinski, 2018). Whereas the holding of elections in authoritarian and hybrid regimes has aggravated the challenge, the problem is not restricted to the Global South or to weakly consolidated democracies, as recent episodes of electoral violence in the United States, Brazil, Turkey, and Hungary remind us. Thus, the promotion of peaceful elections is high on the political agenda for domestic and international agencies alike (Birch and Muchlinski, 2018; Kleinfeld and Sedaca, 2024).

More knowledge about when and where elections turn violent is in high demand across academic and policy-practitioners circles. This study responds to this demand. We build on recent methodological advances in the forecasting literature, along with the burgeoning literature on the causes of electoral violence to predict the likelihood of violence in upcoming elections on a global scale. In recent years, forecasting has gained traction as a complementary approach to understanding political violence and assessing its consequences (e.g. Hegre et al., 2019; Hegre et al., 2021; Vesco et al., 2022; Butcher et al., 2020). Prominent forecasting projects have focused on predicting various forms of political violence, such as the likelihood of future civil wars, fatalities in ongoing wars, and the likelihood of mass atrocities (for a review, see Rød, Gåsste, and Hegre (2024)). However, research on electoral violence has, thus far, primarily been studied through an explanatory framework aimed at

[1] In many non-democratic settings, elections are also held to identify opposition strongholds, signal the strength of the incumbent, provide trappings of democracy, and "divide and rule" tactics (Gandhi and Lust-Okar, 2009).

understanding the circumstances when it occurs and its consequences for democratic processes (e.g. Hafner-Burton, Hyde, and Jablonski, 2013; Fjelde and Höglund, 2014), rather than through a predictive framework that allows policymakers and stakeholders to take proactive steps to minimize its risk and consequences. This study aims to fill this gap by introducing a forecasting system that predicts electoral violence globally. The resulting forecasts can then be used to inform policy decisions and preventive measures to reduce the risk of electoral violence in the future.

Our prediction system uses random forest classifiers trained on historical data on electoral violence alongside a comprehensive set of structural, political, and socioeconomic factors to predict electoral violence on a three-level ordinal scale: no violence, moderate violence, and severe violence. Our predictor features are grouped into thematic constituent models, which are combined for the final forecast using a genetic algorithm. Evaluating our final, weighted, prediction models on historical data between 2014-2023 shows that our model is able to correctly predict the level of electoral violence in 417 and 421 out of the 502 elections in the period when forecasts are made one and two calendar year in ahead of the elections respectively. The models also perform well on other important evaluation metrics such as the Brier score and the AUPR and AUROC metrics.

We then re-train our models using all of the data up until the end of 2023 to generate probabilistic global forecasts of electoral violence in national-level elections for the years 2024 and 2025. These forecasts show an elevated risk for electoral violence primarily in countries with a history of electoral violence and countries with a mixed democratic record, but also in consolidated democracies such as the United States. These forecasts suggest that this tool can provide valuable information for policymakers seeking to mitigate the risk of violence in future elections.

2 Electoral Violence

Electoral violence is a complex phenomenon that can take many forms, ranging from intimidation and harassment to outright violence and coercion (Birch, Daxecker, and Höglund, 2020). Our goal in this paper is to produce forecasts of electoral violence broadly defined, on a global scale. The prediction target must, therefore, satisfy three criteria. First, the target must encompass electoral violence perpetrated by different types of actors. This includes both government-affiliated and opposition-aligned groups. Second, the scope should cover different forms of electoral violence, both lethal and non-lethal physical violence (e.g., beatings, assaults), as well as forms of intimidation that may not involve physical harm but aim to influence the behavior of voters, candidates, or election officials. Third, the target must be coded consistently on a global scale over a sufficiently long time period to allow for the training of machine learning models.

To this end, we have chosen to work with two indicators from the Varieties of Democracy (V-DEM) project (Coppedge et al., 2024b; Pemstein et al., 2024). The first indicator is the *Election Government Intimidation* (v2elintim), measuring the extent to which the government uses intimidation and harassment to influence the outcome of elections. The second indicator is the *Election other electoral violence* (v2elpeace), measuring the extent to which actors other than the government use violence and coercion to influence the outcome of elections (Coppedge et al., 2024a). We have combined these two indicators into a single indicator of the level of electoral violence.

Our indicator for the level of electoral violence is calculated by collapsing the original scale of the two indicators into three categories: *no electoral violence*, for elections where the indicator has the value 3 or higher; *moderate electoral violence*, for elections where the indicator is in the range 1.5-3, and *severe electoral violence* for elections where the indicator is smaller than 1.5. We then calculate the level of electoral violence as the maximum of the two indicators. This allows us to capture the full range of electoral violence, from no violence to severe violence, in a single indicator. If multiple national-level elections take place in a single country year, we take the level of electoral violence to be the maximum across all elections in the country year.

2.1 Prediction target

Based on the definition above, the target for our prediction system is the maximum level of electoral violence observed in each country-year seeing an election. In addition, we limit ourselves to country-years with at least one *national level election*, such as a presidential election or an election to the legislative assembly, and thus exclude country-years which only feature regional elections, referendums, and/or international level elections (e.g. elections to the European parliament). We make this limitation for two main reasons. First, the coding of all election-level data across these types of elections is not complete (Coppedge et al., 2024b) and may vary across countries and contexts, potentially introducing bias in the forecasting system. Second, we believe that these types of elections may have different dynamics compared to the national-level elections, making them less comparable in terms of the drivers and manifestations of electoral violence. Including them in the same forecasting system would thus risk conflating distinct political processes, which could lead to inaccurate predictions or misinterpretation of patterns of violence that are specific to national elections.

To build our forecasting system, we extract data on the level of electoral violence for 1,683 country-years with national-level elections in 172 countries between 1990 and 2023. This includes 779 country-years without electoral violence, 644 with moderate violence and 260 with severe violence. All country-years without elections were excluded from the data set. Forecasts for the future are made at the country-year level, predicting electoral violence one and two years into the future for all countries, regardless of whether a national election is scheduled. This allows us to account for potential unscheduled or early elections, ensuring the model captures potential future risks even in off-cycle years. We trained the models twice, once for elections one year ahead and once for elections two years ahead from the latest available data.

3 Methods

Our forecasting system is built on a set of thematic constituent models, each of which is designed to capture different sets of features that may be relevant for predicting electoral violence. A subset of these models are then combined, using a genetic algorithm, into a weighted ensemble model which produces the final forecast of electoral violence. This approach is in line with the state-of-the-art in conflict forecasting (Hegre et al., 2019; Hegre et al., 2021). In total, we tested 33 different constituent models containing a variety of features, including the history of electoral violence, electoral characteristics, and a wide range of other political, economic, social, and geographic variables. The features are primarily drawn from the Varieties of Democracy (V-DEM) project (Coppedge et al., 2024b), the World Bank's World Development Indicators (WDI) (WorldBank, 2023), and the Digital Society Project (Mechkova et al., 2024). A description of all 33 tested thematic constituent models, including the features included in each, is available in the Appendix (A1).

Predictions were made as probabilities for the three different levels of electoral violence (no, moderate and severe). As our prediction algorithm, we use a standard Random Forest classifier with probability estimates. The Random Forest classifier is a machine learning method that fits a number of decision tree classifiers on random sub-samples of the training data and uses averaging to improve the predictive accuracy and control over-fitting. The Random Forest classifier is commonly used for predicting political violence (Hegre et al., 2019; Hegre et al., 2021; Muchlinski et al., 2016), and has been shown to perform well in a variety of contexts.

Unlike more complex machine learning models, such as deep learning models and gradient-boosted models, the Random Forest model is relatively robust to overfitting and does not require extensive hyperparameter tuning. This makes it a good choice for our forecasting system, as the available training data is relatively limited which can pose a problem for hyperparameter tuning. Initial experiments using a gradient-boosted model showed a high degree of instability in the hyper-parameters, making it difficult to obtain consistent results across different experimental runs. The gradient-boosted model also did not outperform the Random Forest model in terms of predictive performance, which further supported our decision to use the Random Forest model for our forecasting system.

3.1 Training and Evaluation

To properly evaluate the performance of any forecasting system, it is important to ensure that the evaluation is done on data that has not been used when training the model. There are several reasons for this, including the risk of overfitting, the risk of data leakage, and the need to ensure that the model is able to generalize to new data (Ying, 2019; Hernández-Orallo, Flach, and Ferri Ramírez, 2012).

Setting aside a holdout set of data for evaluation is, however, an expensive approach that requires large amounts of data, which may often not be available in practice. As our training data is limited, we instead use a rolling test window approach. In this approach, we iteratively train the model with data up until a certain point, make out-of-sample forecasts for the following time period, then move the training window forward, and repeat the process. This allows us to evaluate the performance of the model on out-of-sample data, while still maximizing the amount of data available for training the model. This evaluation strategy also mimics the real-world forecasting scenario, where the model can be re-trained as new data becomes available (Bergmeir and Benítez, 2012).

Our goal is to make forecasts of electoral violence two years into the future. As the performance of the models may vary across different forecasting horizons, we train the models for the one-year and two-years forecasting horizons separately. We evaluate both horizons in a rolling test window for the period 2014-2023, where the models are trained on data up to one and two years before the forecasted year, respectively. To make the final forecasts, we re-train our models using all available data up until 2023, and make forecasts for 2024 and 2025.

3.2 Performance metrics

To evaluate the performance of the models in our rolling test window, we use four different performance metrics: accuracy, area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPR), and the Brier score.

Accuracy

Accuracy measures the proportion of correctly classified instances, out of all instances in the test set. Accuracy is an intuitive metric that is easy to interpret, but can be misleading when classes are imbalanced, when costs of different types of errors are not equal, or when the difficulty of the classification task varies across different classes.

AUROC

The area under the receiver operating characteristic curve (AUROC) measures the trade-off between the true positive rate, i.e. the proportion of true positives classified as positives, and the false positive rate, i.e. the proportion of true negatives classified as positives, across different thresholds for classifying instances. The AUROC ranges from 0 to 1, where a value of 0.5 indicates that the model performs no better than random, and a value of 1 indicates perfect performance. The AUROC is a useful metric for evaluating the overall performance of a classification model, but can be misleading when the classes are imbalanced (Hernández-Orallo, Flach, and Ferri Ramírez, 2012).

While the AUROC metric was originally designed for binary classification, it can be adapted for multi-class problems using the "one-vs-rest" approach.[2] Here, each class is treated individually as the "positive" class, with the others as "negative." We then average these scores to get the overall AUROC, allowing us to effectively evaluate our multi-class forecasting system (Hernández-Orallo, Flach, and Ferri Ramírez, 2012).

[2]Other alternatives such as class-weighted average AUROC can also be computed.

AUPR

The area under the precision-recall curve (AUPR), or average precision measures the trade-off between precision, i.e. the proportion of true positives among all instances classified as positives, and true positive rate (recall) across different thresholds for classifying instances. The AUPR ranges from 0 to 1, where a value of 0 indicates that the model performs no better than random, and a value of 1 indicates perfect performance. The AUPR is a useful metric for evaluating the performance of a classification model when classes are imbalanced, as it focuses on the positive class (Hernández-Orallo, Flach, and Ferri Ramírez, 2012).

As with the AUROC metric, the AUPR metric was designed for binary classification problems, but can be extended to multi-class classification problems using the one-vs-rest approach.

Brier score

The Brier score is a proper scoring rule that measures the mean squared difference between the predicted probabilities and the actual outcomes. The Brier score ranges from 0 to 1, where a value of 0 indicates perfect performance. The Brier score is a useful metric for evaluating the calibration of a classification model, as it measures the accuracy of the predicted probabilities. The Brier score is particularly useful when the predicted probabilities are used to make decisions, as it directly measures the quality of the predictions (Hernández-Orallo, Flach, and Ferri Ramírez, 2012).

3.3 Ensembling by genetic algorithm

To improve the performance of our forecasting system, we use an ensemble approach to combine the predictions of multiple models. Ensembling is a common technique for improving the performance of machine learning models, as it can help to reduce overfitting, improve generalization, and increase the robustness of predictions (see for instance Montgomery, Hollenbach, and Ward, 2012; Hegre et al., 2019; Rød, Gåsste, and Hegre, 2024). There are many different ways to create ensembles, including naive ensembles, bayesian model averaging, bagging, boosting, and stacking. In this paper, we use a genetic algorithm to optimize the weights of the constituent models in the ensemble (Sivanandam et al., 2008; Holland, 1992).

The genetic algorithm was set up to optimize the Brier score in the rolling test by weighting the constituent models' predictions. Initial weights were set randomly using 1,000 individuals, and the algorithm was run for 1,000 generations with an elitism rate of 0.01, crossover of 0.89, and a kill/replacement-rate of 0.1. In addition, we used a dual-mutation approach with a gene-replacement rate of 0.05 and a gene-mutation rate of 0.1, as well as a regularization threshold such that weights below 0.02 were iteratively set to zero and redistributed among the remaining non-zero weights to force sparsity and model selection in the ensemble. The genetic algorithm was run for the one-year ahead and two-years ahead forecasts separately, generating separate ensembles for each forecast. The final ensemble weights were then used to combine the predictions of the constituent models to make the final forecasts for 2024 and 2025 so that the predicted probability of each election violence level is the weighted average of each constituent model's predicted probability.

3.3.1 Final ensembles for true forecasts

The genetic algorithm identified a total of 12 constituent models with non-zero weights for the final ensembles, with 7 models contributing to the one-year-ahead ensemble and 6 to the two-years-ahead ensemble. The selection of a limited number of models with non-zero weights likely stems from three main factors. First, certain predictor features, particularly those related to the history of electoral violence and irregularities, are highly predictive, leading the genetic algorithm to assign them substantial weight. Second, many constituent models are highly correlated or use similar feature combinations, which causes the algorithm to favor only one or a few models from groups of correlated models. Lastly, the regularization threshold of 0.02, built into the genetic algorithm, imposes sparsity by setting low-weight models to zero.

Table 1 shows the 12 constituent models and their respective weights for the one-year-ahead and two-years-ahead ensembles resulting from the genetic algorithm. More specifically, 7 models contribute to the one-year ahead forecast, and 6 to the two-year ahead forecast. The first group includes models focused on electoral violence history and electoral characteristics, such as irregularities in recent elections. This group accounts for 54% of the weight in the one-year-ahead ensemble and 61% in the two-years-ahead ensemble. The second group comprises broader models related to democracy levels and development, such as those incorporating the full VDEM dataset, combinations of VDEM with WDI and DSP data, and smaller models focusing on mid-level democracy indicators or mixed themes that include democracy indicators and structural features from the WDI. These models make up 38% of the one-year-ahead ensemble and 31% of the two-years-ahead ensemble. The final group consists of models that focus solely on aspects of disinformation, digital security, and social media, particularly from the DSP project, contributing approximately 8% of the weight in both ensembles. A list of the 12 models in the two ensembles, with a brief description of each, can be found in Appendix A1.

Table 1. Weights of the constituent models in the one- and two-year-ahead ensembles, respectively. Ordered by the total weight in either ensemble.

Constituent model	w_{1yr}	w_{2yr}
Election Irregularities last election (short)	0.00	0.47
Election Irregularities last election (long)	0.25	0.14
Election characteristics last election (long)	0.29	0.00
VDEM mid-level indices and WDI structural	0.00	0.13
VDEM civil liberties indicies	0.00	0.11
VDEM mid-level indices, WDI structural, and DSP infrastructure	0.10	0.00
Full VDEM, WDI, and DSP model	0.10	0.00
Full VDEM model	0.09	0.00
VDEM mid-level indices (alternative)	0.09	0.00
DSP disinformation, social media climate and usage	0.00	0.08
Full DSP model	0.08	0.00
VDEM accountability indices	0.00	0.07

4 Results

4.1 Out-of-sample evaluation

The performance of the models was evaluated on all elections between 2014 and 2023, using a rolling test window approach, in which the models were retrained each year on data up to the previous year. The evaluation was based on our four evaluation metrics: accuracy, AUROC, AUPR, and the Brier score, which all capture different aspects of the model's performance.

The predictive performance of the final ensemble models is shown in Figure 1. The ensemble models show high performance across all evaluation metrics, with accuracy levels of $\geq$ 80% and AUPR $\geq$ 0.8 and AUROC $\geq$ 0.9 indicating that they do very well in distinguishing between positive and negative classes across different classification thresholds. Additionally, Brier scores around 0.1 are furthermore indicative of well-calibrated predictions, highlighting that the models are not only able to distinguish well between the classes, but do so with high sharpness. The corresponding evaluation metrics for all 33 thematic constituent models can be found in Appendix B. Importantly, the ensemble models perform on par with or better than the top constituent models on all evaluation metrics, suggesting that the ensemble predictions are more robust and reliable than the individual constituent models.

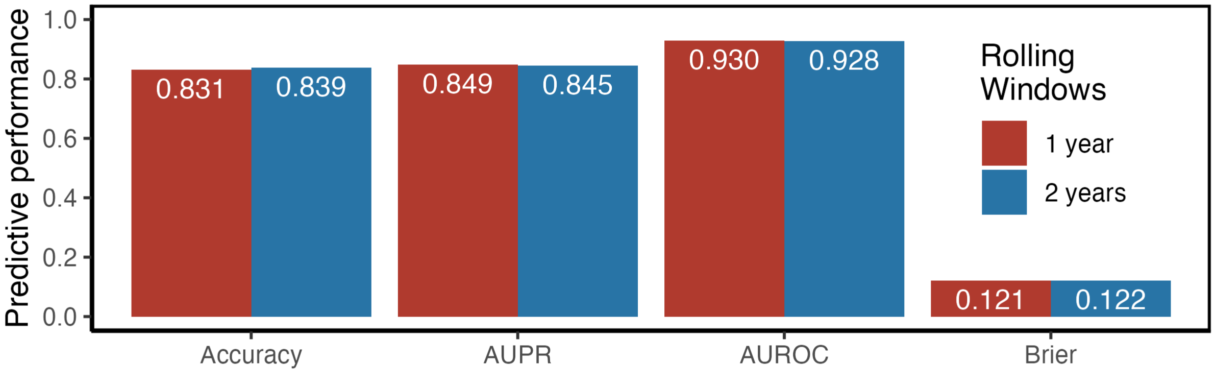
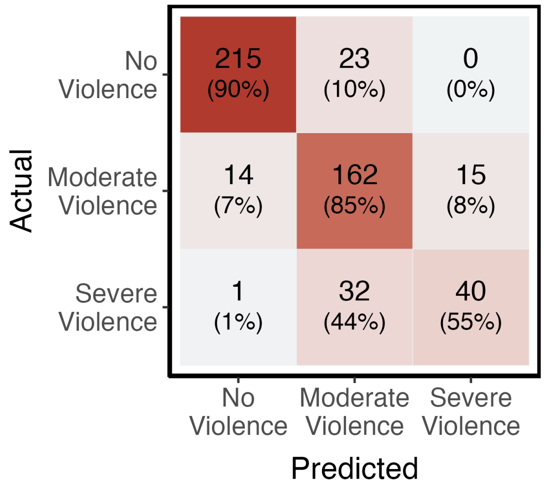


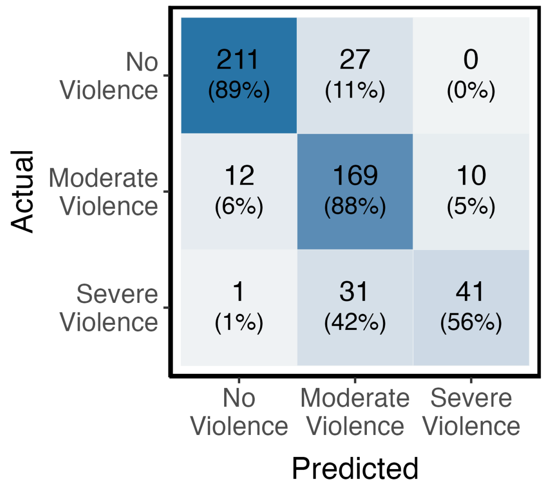
Figure 1. Evaluation metrics: Accuracy, AUPR, and AUROC for 502 country years with elections across the 2014–23 period, for models trained on data up to one year prior to the evaluation year (1-year rolling window) or two years prior (2-year rolling window)

Figure 2 shows the observed level of electoral violence against the predicted level across the 502 country-years with elections over the 2014–23 period.[3] One year into the future, the model predicted 55 cases to have severe electoral violence. Of these 55, 40 did, in fact, experience severe violence, while 15 experienced a moderate level of violence. None of the 55 cases predicted to have severe violence experienced no electoral violence. Thus, there are only a few instances where the model exaggerates the risk of violence. The rate of false alarms is also low: one year into the future, the model predicted 230 out of the 502 country years to have *no violence*. Only 7% of these predictions were incorrect. Of the 73 elections that did experience severe violence in the period, the model correctly predicts 55% of cases. In only one case where severe electoral violence occurred, in Benin 2019, did the model predict no violence as the likeliest outcome. In this case, however, the model gives the no and moderate levels of violence an almost equal probability around 48%.

[3]'Predicted level' here means the election violence level with the largest probability according to the ensemble model.



(a) 1 year rolling window



(b) 2 years rolling window

Figure 2. Confusion matrices: Actual outcome versus predicted outcome for 502 country years with elections across the 2014–23 period, for models trained on data up to one year prior to the evaluation year (1-year rolling window) or two years prior (2-year rolling window).

The results of the out-of-sample evaluation in the rolling test window show that the final ensemble models demonstrate strong predictive accuracy, effectively distinguishing between classes with high reliability and well-calibrated probabilities. They also outperform, or match, the individual thematic models, indicating that the predictions are robust. The low overall error rates and the near-zero misclassification between no and severe electoral violence further highlight the reliability of the model. This overall performance gives confidence in the model's ability to make accurate forecasts for the future.

4.2 Forecasts for 2024–2025

Using the final ensemble of constituent models, we forecast the probability of electoral violence for the years 2024 and 2025. The forecasts from our model are shown in Figure 3 below, which displays the likelihood of *any* level of electoral violence – either moderate or severe.[4] Countries with a low predicted risk of experiencing *any* level of electoral violence are shown in blue, countries at a medium predicted risk of experiencing *any* level of electoral violence are shown in yellow, and countries at a high risk of experiencing *any* level of electoral violence are shown in red.

[4]Detailed predictions for all countries, regardless of whether they have scheduled elections in 2024 and 2025, can be found at the Kofi Annan Foundation's Electoral Vulnerability Index website or can be made available by the corresponding author upon request.

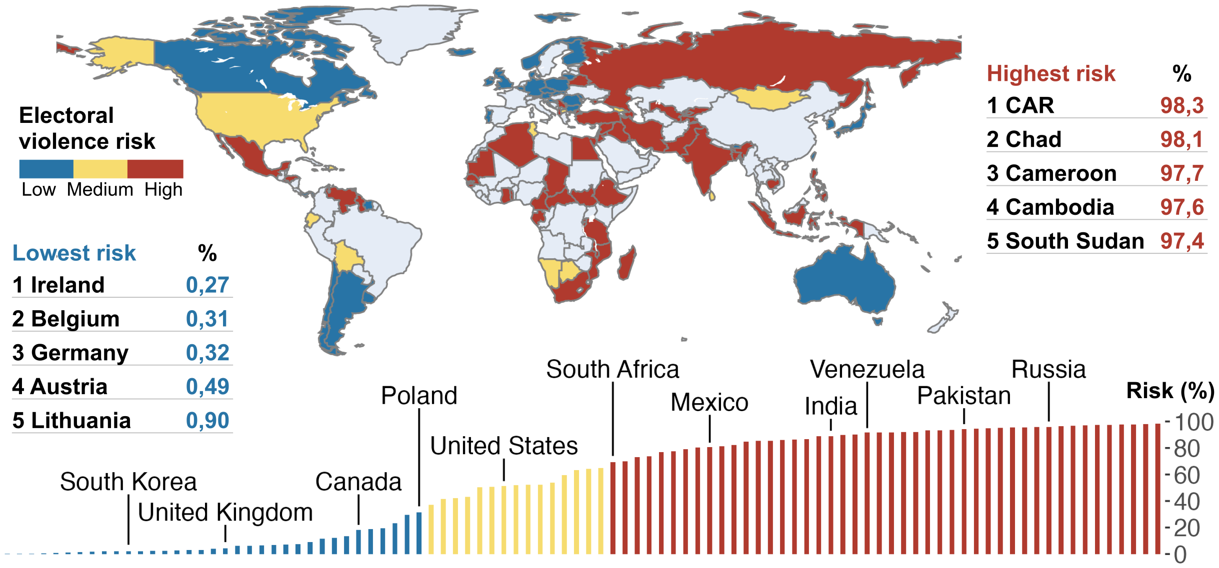


Figure 3. Predicted probability for the risk of either moderate or severe election violence in 2024 or 2025. Countries without scheduled elections are marked with gray shade.

The forecasts show that the likelihood of electoral violence is highest among countries with a history of violence or intimidation around elections, low levels of democracy, and high levels of corruption. For example, the model predicts a high likelihood of electoral violence in countries such as Equatorial Guinea, Egypt, and Chad. In contrast, countries with high levels of democracy and low levels of corruption, such as Denmark, Austria, and Ireland, are predicted to have a low likelihood of electoral violence, unsurpringly. However, while both the highest predicted risk and the highest level of electoral violence intensity are forecasted in countries with low levels of democracy, this does not mean that this is a problem confined to non-consolidated democracies. Countries with long democratic traditions, including the United States, Botswana, and India also have elevated risks of experiencing any level of electoral violence. Moreover, it is important to note that the definition of electoral violence includes intimidation and harassment, which may not necessarily be expressed as overt, physical violence. Indeed, when repression is sufficiently severe, physical violence might not occur, even though the electoral environment is highly coercive. This caveat is important when interpreting the forecasts for countries with high levels of political repression.

4.3 Case examples, 2024

Our models are trained on data up until 2023 and forecasts are made as true out-of-sample predictions for 2024 and 2025. To evaluate the predictions for 2024, we compare our forecasts with qualitative evidence from electoral processes already concluded or ongoing at the time of writing. We focus on three countries: Venezuela, India, and the United States. These countries were chosen as they have experienced varying levels and types of electoral violence in the past, and their elections in 2024 represent cases with medium to high predicted risk of any level of electoral violence. We present our forecasts for these countries, as well as a brief summary of the electoral violence that has taken place in the context of these elections. While the final country ratings of electoral violence for the 2024 elections in these cases are not yet included in the data we rely on to train our prediction model, case evidence suggests that these countries will most likely be coded as having at least a moderate level of electoral violence. We believe that these cases offer early validation of the predictive performance of our model.

4.3.1 Venezuela

The Venezuelan presidential election was held on July 27th 2024, and our model predicted a high likelihood of electoral violence for this election, with an approximate 55% probability of limited violence and 38% probability of severe violence.

In the lead-up to the election, the government engaged in widespread repression. Opposition candidates were arbitrarily disqualified, including prominent figures such as the politician María Corina Machado, and activists were arrested or harassed. A month ahead of the election, at least 76 arbitrary detentions were documented during the campaign period (CEPAZ, 2024; Human Rights Watch, 2024b). The result of the election was widely contested, with both the opposition and government claiming victory, leading to widespread protests and violence. The UN Human Rights Council's Fact-Finding Mission in Venezuela recorded at least 23 deaths between July 28 and August 8. Furthermore, over 1260 people were detained in conjuncture with the election, with serious due process violations such as remote hearings, unjust charges, and restricted access to legal counsel (OHCHR, 2024)

4.3.2 India

The Indian general election took place between April 19th and June 1st 2024. Our model predicted a high likelihood of electoral violence for this election, with an approximate 75% probability of limited violence and 14% probability of severe violence.

The pre-election and election periods were marked by widespread violence and intimidation in several parts of India. This violence was both politically and communally charged, involving various actors, but mainly including supporters of the ruling Bharatiya Janata Party (BJP) party and state agencies controlled by them, along with opposition parties. The electoral landscape in India has been shaped by physical altercations, targeted assassinations, and systemic suppression, particularly impacting minority groups such as Muslims and Christians Human Rights Watch, 2024a; Amnesty International, 2024

One of the most notable examples comes from West Bengal, where tensions between the BJP and the Trinamool Congress (TMC) have resulted in repeated violent clashes. Violent encounters, including the use of crude bombs, stone-pelting, and car blockades, have been reported. Notably, in April 2024, more than 100 complaints of election-related violence were filed by both parties, illustrating the volatile atmosphere (The London Story, 2024).

4.3.3 United States

The United States presidential and general elections will take place on November 5th 2024. Our model predicts a relatively high likelihood of electoral violence for this election, with an approximate 51% probability of limited violence and 1% probability of severe violence.

At the time of writing, the election cycle in the U.S. has been shaped by widespread intimidation and several significant instances of overt violence. Presidential candidates have been at the forefront of election-related violence and intimidation. Former President Donald Trump has faced two assassination attempts(Associated Press, 2024b). Another alleged attempt was thwarted in Palm Beach, Florida (Reuters, 2024a). Democratic candidates are also under threat. Vice President Kamala Harris and President Biden were targeted with violent online threats (Reuters, 2024b), further fuelling tensions. Election officials have also faced increased harassment and intimidation since the 2020 election, reflecting broader trends of rising threats against those tasked with upholding electoral integrity (Brennan Center, 2024). Recent threats include suspicious packages sent to election officials in 15 states, some containing hazardous substances such as fentanyl (Associated Press, 2024a).

5 Discussion

Forecasting of political violence has in recent years been propelled into the mainstream of conflict research, with a growing number of ambitious projects aiming to predict the likelihood of violence across various contexts (e.g. Hegre et al., 2019; Hegre et al., 2021; Mueller and Rauh, 2018; Vesco et al., 2022; Rød, Gåsste, and Hegre, 2024). The prediction of electoral violence specifically is a challenging task due to the complexity of the phenomenon and the multitude of factors that can influence its occurrence (Birch, Daxecker, and Höglund, 2020; Höglund, 2009; Birch and Muchlinski, 2018). Our forecasting system is built on thematic constituent models and ensembling, similar to state-of-the-art conflict forecasting systems (Hegre et al., 2019; Hegre et al., 2021).

The evaluation of our model on historical data shows that our forecasting system can provide valuable insights into the likelihood of, and therefore prevention of, electoral violence. The genetically weighted ensemble performs particularly well on the performance metrics in out-of-sample evaluation in 2014-2023. This type of forward-looking predictions can be of great value to policymakers, election observers, and other stakeholders, as they can help to identify countries at risk of violence and assist in implementing preventive measures to mitigate the risk.

In this paper, we use a broad definition of electoral violence, including intimidation and harassment as forms of violence. This is in line with the literature on electoral violence, which often includes a wide range of behaviors that can be seen as violent or coercive (Birch, Daxecker, and Höglund, 2020). However, it is important to note that our forecasting system does not distinguish between different types of violence, only the level of violence. Because intimidation and harassment are included in our definition of electoral violence, highly repressive electoral environments where no physical violence is observed may still be classified as suffering from severe electoral violence. Therefore, countries with high levels of predicted risk of electoral violence may not necessarily experience widespread overt, physical violence. Notably, severe intimidation and coercion can be used in lieu of physical violence. However, even in highly repressive countries, a high predicted risk of electoral violence may still be a valuable signal for policymakers and other stakeholders, as it can help to identify countries where the risk of violence is particularly high, e.g. if the regime is challenged after the election, or if demonstrations or other forms of protest arise in conjuncture with the election.

Our true out-of-sample forecasts for 2024 and 2025 identify countries of particular risk for electoral violence. The countries with the highest predicted risk and level of electoral violence are countries with a long history of electoral violence and which are not considered fully democratic.

In these countries, the regime is likely to use violence to maintain power either by subverting the electoral process using harassment and intimidation ahead of the election or through explicit physical violence following the election. Examples of countries in this category are Belarus, Venezuela, and Zimbabwe. However, our forecasts also identify countries with a high level of predicted risk that have considerably longer democratic traditions, such as Ghana, India, and Poland. In these countries, the risk is primarily for less severe levels of violence, but it still highlights the potential for violence even in consolidated democracies. Of particular note is also our prediction for electoral violence in the United States which according to our forecasting system has a 51% likelihood of experiencing electoral violence in the 2024 general election. This is the highest probability among all countries classified as liberal democracies by VDEM (Coppedge et al., 2024a).

In conclusion, our study demonstrates that it is possible to build a forecasting system for electoral violence with a high degree of predictive accuracy using state-of-the-art machine learning tools. By identifying countries at high risk, this forecasting system offers valuable insights for policymakers and election observers seeking to prevent violence and promote democratic integrity. Future work could expand the scope of these models to include more granular data and explore ways to integrate additional factors that may influence electoral violence. While challenges remain, particularly in distinguishing between types of violence, this study represents an important step toward more effective forecasting of electoral violence globally.

Acknowledgements

The research was funded by the Kofi Annan Foundation Electoral Vulnerability Index (EVI) project, the European Research Council, project H2020-ERC-2015-AdG 694640 (ViEWS), Riksbankens Jubileumsfond, grant M21-0002 (Societies at Risk), the European Research Council under Horizon Europe (Grant agreement No. 101055176, ANTICIPATE), Norwegian Research Council (NFR) grant 334977 (The Uncertainty of Forecasting Fatalities in Armed Conflict (UFFAC)), Knut and Alice Wallenberg foundation (2017.0141), the The Royal Swedish Academy of Letters, History and Antiquities, and the Center for Advanced Study (CAS) at the Norwegian Academy of Science and Letters.

References

Amnesty International (2024). *India: Crackdown on opposition reaches a crisis point ahead of national elections*. https://www.amnesty.org/en/latest/news/2024/03/india-crackdown-on-opposition-reaches-a-crisis-point-ahead-of-national-elections. [Accessed 20-09-2024].

Associated Press (2024a). *The FBI is investigating suspicious packages sent to election officials in more than 15 states*. https://apnews.com/article/elections-workers-security-suspicious-packages-e3400b1e86bc02f7345d9970ef356bec. [Accessed 30-09-2024].

— (2024b). *Trump injured but 'fine' after attempted assassination at rally, shooter and one attendee are dead*. https://apnews.com/article/trump-vp-vance-rubio-7c7ba6b99b5f38d2d840ed9 [Accessed 17-10-2024].

Bergmeir, Christoph and José M Benítez (2012). "On the use of cross-validation for time series predictor evaluation". In: *Information Sciences* 191, pp. 192–213.

Birch, Sarah, Ursula Daxecker, and Kristine Höglund (2020). *Electoral violence: An introduction*.

Birch, Sarah and David Muchlinski (2018). "Electoral violence prevention: what works?" In: *Democratization* 25.3, pp. 385–403.

Brennan Center (2024). *Local Election Officials Survey*. https://www.brennancenter.org/our-work/research-reports/local-election-officials-survey-april-2023. [Accessed 01-10-2024].

Butcher, Charles, Benjamin E Goldsmith, Sascha Nanlohy, Arcot Sowmya, and David Muchlinski (2020). "Introducing the targeted mass killing data set for the study and forecasting of mass atrocities". In: *Journal of Conflict Resolution* 64.7-8, pp. 1524–1547.

CEPAZ (2024). *REC: En Venezuela ha aumentado la violencia política en el contexto electoral*. cepaz.org/rec-en-venezuela-ha-aumentado-la-violencia-politica-en-el-contexto-electoral. [Accessed 17-09-2024].

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Fabio Angiolillo, Michael Bernhard, Cecilia Borella, Agnes Cornell, M. Steven Fish, Linnea Fox, Lisa Gastaldi, Haakon Gjerløw, Adam Glynn, Ana Good God, Sandra Grahn, Allen Hicken, Katrin Kinzelbach, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Anja Neundorf, Pamela Paxton, Daniel Pemstein, Johannes Rydén Oskar abd von Römer, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Aksel Sundström, Eitan Tzelgov, Luca Uberti, Yi-ting Wang, Tore Wig, and Daniel Ziblatt (2024a). "V-Dem Codebook v14". In: *Varieties of Democracy (V-Dem) Project*.

Coppedge, Michael et al. (2024b). *V-Dem [Country-Year/Country-Date] Dataset v14*. URL: https://doi.org/10.23696/mcwt-fr58.

Fjelde, Hanne and Kristine Höglund (Dec. 2014). "Electoral Institutions and Electoral Violence in Sub-Saharan Africa". In: *British Journal of Political Science* FirstView.02, pp. 1–24. URL: http://journals.cambridge.org/article{\%}5C{_}S0007123414000179papers3://publication/doi/10.1017/S0007123414000179.

Gandhi, Jennifer and Ellen Lust-Okar (June 2009). "Elections Under Authoritarianism". In: *Annual Review of Political Science* 12.1, pp. 403–422. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev.polisci.11.060106.095434papers3://publication/doi/10.1146/annurev.polisci.11.060106.095434.

Hafner-Burton, Emilie M, Susan D Hyde, and Ryan S Jablonski (2013). "When Do Governments Resort to Election Violence?" In: *British Journal of Political Science* 44, pp. 149–

179. DOI: 10.1017/S0007123412000671. URL: http://dx.doi.org/doi:10.1017/S0007123412000671..

Hegre, Håvard, Marie Allansson, Matthias Basedau, Michael Colaresi, Mihai Croicu, Hanne Fjelde, Frederick Hoyles, Lisa Hultman, Stina Högbladh, Remco Jansen, et al. (2019). "ViEWS: A political violence early-warning system". In: *Journal of peace research* 56.2, pp. 155–174.

Hegre, Håvard, Curtis Bell, Michael Colaresi, Mihai Croicu, Frederick Hoyles, Remco Jansen, Maxine Ria Leis, Angelica Lindqvist-McGowan, David Randahl, Espen Geelmuyden Rød, et al. (2021). "ViEWS2020: revising and evaluating the ViEWS political violence early-warning system". In: *Journal of peace research* 58.3, pp. 599–611.

Hernández-Orallo, José, Peter Flach, and César Ferri Ramírez (2012). "A unified view of performance metrics: Translating threshold choice into expected classification loss". In: *Journal of Machine Learning Research* 13, pp. 2813–2869.

Höglund, Kristine (2009). "Electoral violence in conflict-ridden societies: Concepts, causes, and consequences". In: *Terrorism and political violence* 21.3, pp. 412–427.

Holland, John H (1992). "Genetic algorithms". In: *Scientific american* 267.1, pp. 66–73.

Human Rights Watch (2024a). *India: Hate Speech Fueled Modi's Election Campaign*. https://www.hrw.org/news/2024/08/14/india-hate-speech-fueled-modis-election-campaign. [Accessed 20-09-2024].

— (2024b). *Venezuela: Brutal Crackdown on Protesters, Voters*. https://www.hrw.org/news/2024/09/04/venezuela-brutal-crackdown-protesters-voters. [Accessed 17-10-2024].

Hyde, Susan D. and Nikolay Marinov (2012). "Which Elections Can Be Lost?" In: *Political Analysis* 20.2, pp. 191–210. ISSN: 10471987. DOI: 10.1093/pan/mpr040.

Kleinfeld, Rachel and Nicole Bibbins Sedaca (2024). "How to prevent political violence". In: *Journal of Democracy* 35.4, pp. 35–45.

Mechkova, Valeriya, Daniel Pemstein, Brigitte Seim, and Steven Wilson (2024). *Digital Society Project Dataset v6*.

Montgomery, Jacob M, Florian M Hollenbach, and Michael D Ward (2012). "Improving predictions using ensemble Bayesian model averaging". In: *Political Analysis* 20.3, pp. 271–291.

Muchlinski, David, David Siroky, Jingrui He, and Matthew Kocher (2016). "Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data". In: *Political Analysis* 24.1, pp. 87–103.

Mueller, Hannes and Christopher Rauh (2018). "Reading between the lines: Prediction of political violence using newspaper text". In: *American Political Science Review* 112.2, pp. 358–375.

OHCHR (2024). *Venezuela Fact-Finding Mission calls for end of repression, thorough investigations, after elections*. https://www.ohchr.org/en/press-releases/2024/08/venezuela-fact-finding-mission-calls-end-repression-thorough-investigations. [Accessed 01-10-2024].

Pemstein, Daniel, Kyle L. Marquardt, Eitan Tzelgov, Yi-ting Wang, Juraj Medzihorsky, Joshua Krusell, Farhad Miri, and Johannes von Römer (2024). *The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data*. V-Dem Working Paper No. 21, 9th edition. University of Gothenburg: Varieties of Democracy Institute.

Reuters (2024a). *Arizona man wanted for threatening to kill Trump*. https://www.reuters.com/world/us/arizona-man-wanted-threatening-kill-trump-2024-08-22/. [Accessed 17-10-2024].

Reuters (2024b). *Virginia man charged with threatening Vice President Harris*. https://www.reuters.com/world/us/virginia-man-charged-with-threatening-vice-president-harris-2024-08-05/. [Accessed 01-10-2024].

Rød, Espen Geelmuyden, Tim Gåsste, and Håvard Hegre (2024). "A review and comparison of conflict early warning systems". In: *International Journal of Forecasting* 40.1, pp. 96–112.

Sivanandam, SN, SN Deepa, SN Sivanandam, and SN Deepa (2008). *Genetic algorithms*. Springer.

The London Story (2024). *Tech Harms and Human Rights in India's 2024 Elections*. https://thelondonstory.org/wp-content/uploads/Tech-Harms-and-Human-Rights-in-Indias-2024-Elections.pdf. [Accessed 20-09-2024].

Vesco, Paola, Håvard Hegre, Michael Colaresi, Remco Bastiaan Jansen, Adeline Lo, Gregor Reisch, and Nils B Weidmann (2022). "United they stand: Findings from an escalation prediction competition". In: *International Interactions* 48.4, pp. 860–896.

WorldBank (2023). *World Development Indicators*. Washington DC: World Bank.

Ying, Xue (2019). "An overview of overfitting and its solutions". In: *Journal of physics: Conference series*. Vol. 1168. IOP Publishing, p. 022022.

Appendix A1: Constituent models description for models in the final ensembles

A short description of the models included in the final ensembles, ordered by their total weight, is included below. The exact features included can be found in Appendix A2.

1. Election Irregularities last election (short): All Varieties of democracy (VDEM) Country-Date features pertaining to election irregularities, including the level of electoral violene perpetrated by government and non-government actors, in the last held election in the training data.

2. Election Irregularities last election (long): All features from (1), as well as features measuring the number of elections in a row that have seen no/moderate/severe electoral violence, and the number of elections since the last major constitutional change.

3. Election characteristics last election (long): All VDEM Country-Date features pertaining to the election characteristics (including irregularities) in the last held election in the training data, as well as the streak-variables from (2).

4. VDEM mid level indicies and WDI structural: VDEM Coutry-Year mid level indicators of democracy, such as freedom of expression and associacion, share of population with sufferace etc, as well as structural features from the World Development Indicators (WDI) such as GDP, population, and infant moratlity.

5. VDEM civil liberties: All VDEM Country-Year features relating to civil liberties.

6. VDEM mid level indicies, WDI structural, and DSP infrastructure: Same features as (4) but also including features from the Digital Society Project (DSP) pertaining to digital infrastructure.

7. Full VDEM, WDI, and DSP model: All features across the VDEM, WDI, and DSP datasets.

8. Full VDEM model: All features across in the VDEM datasets.

9. VDEM mid level indicies (alternative): VDEM Coutry-Year mid level indicators of democracy, such as civil liberties, neopatrimonialism, corruption, and gender equality.

10. DSP disinformation, social media climate and usage: Features from the DSP pertaining to desinformation, social media usage, and social media climate, including government and political parties dissemination of false information.

11. Full DSP model: All interval-scale features from the DSP project.

12. VDEM accountability indicvies: All VDEM accountability indicies, such as the horizontal, vertical, and diagonal accountability indicies.

Appendix A2: Constituent models description for all models

Below are tables describing the 33 thematic constituent models evaluated for the final ensemble, divided into five categories: 1) constituent models focusing on the characteristics of the previous election; 2) constituent models using the Digital Society Project indicators for digital infrastructure and vulnerability; 3) constituent models focusing on VDEM yearly indicators; 4) constituent models using the World Development Indicators; and 5) combination models which mix features across the four data sources.

Table 2. Election history (VDEM-CD) constituent models

Model name	**Description of features**	**Included features**
History of electoral violence (history only)	Features tracking streaks of peaceful, severely violent, and low-violence elections, and # elections since the last constitutional change	cons_elect, peaceful_streak, violent_streak, lowviolent_streak
History of electoral violence (full)	Features from history of electoral violence model, and reported levels of electoral violence in the last election	"History of electoral violence (history only)" plus v2elintim_osp, v2elpeace_osp
Election Irregularities last election (short)	Irregularity-related features from the last election	v2elembaut, v2elembcap, v2elmulpar, v2elvotbuy, v2elirreg, v2elintim, v2elpeace, v2elboycot, v2elfrfair, v2elmonden, v2elmonref
Election Irregularities last election (long)	Irregularity-related features from the last election, including violence streaks	"Election Irregularities last election (short)" plus cons_elect, peaceful_streak, violent_streak, lowviolent_streak
Election Characteristics last election (structural)	Structural features from the last election	v2asuffrage, v2elcomvot, v2elgvsuflvl, v2eldonate, v2elpubfin, v2elembaut, v2elembcap, v2elmulpar, v2elrgstry, v2elvotbuy, v2elfrcamp, v2elpdcamp, v2elpaidig, v2eldommon, v2elintmon, v2elvaptrn
Election Characteristics last election	Characteristics of the last election features from the V-Dem country-date dataset	"Election Characteristics last election (structural)" plus v2elirreg, v2elintim, v2elpeace, v2elboycot, v2elfrfair, v2elmonden, v2elmonref, v2elaccept, v2elasmoff, cons_elect
Election Characteristics last election (long)	Characteristics of the last election features from the V-Dem country-date dataset, including violence streaks	"Election Characteristics last election" plus peaceful_streak, violent_streak, lowviolent_streak

Table 3. Digital Society Project constituent models

Model name	Description of features	Included features
DSP Monitoring	DSP features relating to government monitoring, surveillance, and repression online	v2smregcap, v2smgovfilprc, v2smgovsmmon, v2smgovsmcenprc, v2smarrest
DSP Disinformation and social media climate and usage	DSP features relating to disinformation online and social media usage	v2smgovdom, v2smpardom, v2smfordom, v2smorgelitact, v2smcamp
DSP Social media climate	Social media climate from DSP, including dissemination of disinformation, online polarization and hate speech, and traditional use of social media by elites/political candidates	"DSP Disinformation and social media usage" plus v2smonper, v2smmefra, v2smpolsoc, v2smpolhate
DSP Social Media Climate, security	DSP features relating to social media climate, security, and usage	"DSP Social media climate" plus v2smgovcapsec, v2smpolcap
DSP Infra	Digital infrastructure features, including media features from V-Dem-CY, embassy capacity from V-Dem-CY, cyber security + monitoring and surveillance of social media from DSP, and internet use from WDI	"DSP Monitoring" plus v2smonex, v2elfrcamp, v2mecrit, v2merange, v2elembaut, it.net.user.zs
DSP full model	All interval scale features from DSP	"DSP Disinformation and social climate and usage" plus v2smgovab, v2smparab, v2smforads, v2smgovfilcap, v2smgovshutcap, v2smgovshut, v2smgovsm, v2smgovsmalt, v2smgovcapsec, v2smregcon, v2smprivex, v2smprivcon, v2smregapp, v2smlawpr, v2smdefabu, v2smonex, v2smorgviol, v2smorgavgact

Table 4. VDEM Country-Year constituent models

Model name	Description of features	Included features
VDEM Political Exclusion Indicies	VDEM-CY features on exclusion of groups	v2xpe_exlecon, v2xpe_exlgender, v2xpe_exlgeo, v2xpe_exlpol, v2xpe_exlsocgr
VDEM Neopatrimonialism	VDEM-CY neopatrimonialism features	v2x_neopat, v2xnp_client, v2xnp_pres, v2xnp_regcorr
VDEM Civil Liberties Indicies	VDEM-CY features on civil liberties	v2x_clphy, v2x_clpol, v2x_clpriv, v2x_civlib
VDEM Accountability Indicies	VDEM-CY features on accountability	v2x_accountability, v2x_veracc, v2x_diagacc, v2x_horacc
VDEM Gender	VDEM-CY gender features	v2x_gencl, v2x_gencs, v2x_genpp, v2x_gender
VDEM High level indicies	VDEM-CY high-level indices	v2x_polyarchy, v2x_libdem, v2x_partipdem, v2x_delibdem, v2x_egaldem
VDEM mid level indicies (alternative)	VDEM-CY mid-level indices	"VDEM Accountability Indicies" plus v2x_neopat, v2x_civlib, v2x_gender, v2x_corr, v2x_rule, v2xcs_ccsi, v2xps_party, v2x_divparctrl, v2x_feduni
VDEM mid level indicies	VDEM-CY and CD mid-level component indices	"VDEM High level indicies" plus v2x_api, v2x_mpi, v2x_freexp_altinf, v2x_frassoc_thick, v2x_suffr, v2xel_frefair, v2x_elecoff, v2xcl_rol, v2x_jucon, v2xlg_legcon, v2x_cspart, v2xdd_dd, v2xel_locelec, v2xel_regelec, v2xdl_delib, v2xeg_eqprotec, v2xeg_eqaccess
VDEM full model	All v2x indices from VDEM-CY	"VDEM mid level indicies" plus v2x_ex_confidence, v2x_ex_direlect, v2x_ex_hereditary, v2x_ex_military, v2x_ex_party, v2xnp_client, v2xnp_pres, v2xnp_regcorr, v2xdd_cic, v2xdd_i_ci, v2xdd_i_rf, v2xdd_toc, v2xdd_i_pl, v2xdd_i_or, v2xcs_ccsi, v2x_EDcomp_thick, v2xcl_disc, v2xcl_dmove, v2xcl_slave, v2xex_elecleg, v2xme_altinf, v2xps_party, v2x_divparctrl, v2x_feduni, v2xca_academ

Table 5. World Development Indicators Constituent model

Model name	Description of features	Included features
WDI Education	WDI education factors, including enrollment and expenditure	se.enr.prim.fm.zs, se.enr.prsc.fm.zs, se.prm.nenr, se.xpd.totl.gb.zs, se.xpd.totl.gd.zs
WDI Resources	WDI factors on resources and GDP	ny.gdp.pcap.kd, ny.gdp.pcap.kd.zg, dt.oda.odat.pc.zs, ny.gdp.petr.rt.zs, ny.gdp.totl.rt.zs
WDI Structural	WDI structural factors, including population, age composition, IMR, life expectancy, and GDP	sp.pop.totl, ny.gdp.pcap.kd, ny.gdp.pcap.kd.zg, sp.dyn.imrt.in, sp.dyn.le00.in, sp.pop.0014.fe.zs, sp.pop.grow, sp.pop.65up.fe.zs
WDI full	WDI full model	"WDI Education", "WDI Resources", and "WDI Structural" plus ms.mil.xpnd.zs, ms.mil.xpnd.gd.zs, nv.agr.totl.kn, sp.dyn.le00.in, sh.sta.maln.zs, sh.sta.stnt.zs, sl.tlf.totl.fe.zs, sm.pop.totl.zs, sh.dyn.mort.fe, sp.pop.1564.fe.zs, sp.urb.totl.in.zs, sl.uem.neet.zs, it.net.user.zs

Table 6. Combination Constituent models

Model name	Description of features	Included features
VDEM High level indicies and WDI structural	Combination of features from VDEM High level indices and WDI structural	”VDEM High level indicies” and ”WDI Structural”
VDEM Mid level indicies and WDI structural	Combination of features from VDEM mid level indices and WDI structural	”VDEM mid level indicies” and ”WDI Structural”
Election Irregularities (last election), VDEM civil liberties, and WDI structural	Combination of features from VDEM Civil Liberties, election irregularities (last election), and WDI structural	”VDEM Civil Liberties Indicies”, ”Election Irregularities last election (short)”, and ”WDI Structural”
Election Irregularities (last election), VDEM exclusion, and WDI structural	Combination of features from VDEM Political Exclusion, election irregularities (last election), and WDI structural	”VDEM Political Exclusion Indicies”, ”Election Irregularities last election (short)”, and ”WDI Structural”
VDEM Mid level indicies, WDI structural, and DSP infrastructure	Combination of features from VDEM mid level indices, WDI structural, and DSP infrastructure	”VDEM Mid level indicies and WDI structural” and ”DSP Infra”
Full VDEM, WDI, and DSP model	Combination of all features above	”VDEM Mid level indicies, WDI structural, and DSP infrastructure”, ”Election Irregularities (last election), VDEM civil liberties, and WDI structural”, and ”Election Irregularities (last election), VDEM exclusion, and WDI structural”

Appendix B: Model Performance

Table 7. Performance of models in the one year ahead prediction task

Rank	Model	Accuracy	Brier	AUROC	AUPR
1	Genetically optimized ensemble	0.831	0.121	0.930	0.849
2	Election Irregularities (last election), VDEM exclusion and WDI structural	0.843	0.123	0.905	0.807
3	Election Characteristics last election (full)	0.834	0.124	0.921	0.839
4	Election Irregularities (last election), VDEM civil liberties, and WDI structural	0.850	0.125	0.907	0.812
5	Election Irregularities last election (long)	0.831	0.126	0.920	0.833
6	Election Irregularities last election (short)	0.833	0.126	0.920	0.835
7	Election Characteristics last election (full)	0.836	0.126	0.919	0.837
8	VDEM Mid level indicies, WDI structural, and DSP infrastructure	0.833	0.129	0.914	0.815
9	History of electoral violence (full)	0.829	0.131	0.914	0.821
10	VDEM Mid level indicies and WDI structural	0.824	0.131	0.909	0.805
11	VDEM full model	0.829	0.132	0.919	0.832
12	VDEM mid level indicies	0.817	0.138	0.916	0.841
13	VDEM High level indivies and WDI structural	0.824	0.139	0.903	0.807
14	DSP full model	0.810	0.139	0.917	0.809
15	VDEM mid level indicies	0.805	0.145	0.902	0.802
16	DSP Social Media Climate, security	0.801	0.150	0.905	0.804
17	Election Characteristics last election, structural	0.798	0.151	0.895	0.806
18	DSP Infra	0.788	0.151	0.889	0.780
19	DSP Disinformation and social media usage	0.801	0.156	0.892	0.786
20	DSP Disinformation and social climate and usage	0.797	0.156	0.892	0.787
21	VDEM High level indicies	0.781	0.157	0.891	0.788
22	DSP Social media climate	0.801	0.157	0.891	0.784
23	VDEM Political Exclusion Indicies	0.781	0.158	0.893	0.797
24	History of electoral violence (history only)	0.779	0.160	0.883	0.763
25	Full model (all features)	0.781	0.166	0.885	0.781
26	VDEM Neopatrimonialism	0.747	0.171	0.864	0.750
27	WDI Structural	0.743	0.174	0.853	0.701
28	DSP Monitoring	0.761	0.176	0.875	0.746
29	VDEM Accountability Indicies	0.751	0.176	0.862	0.760
30	VDEM Civil Liberties Indicies	0.735	0.186	0.850	0.734
31	WDI full model	0.758	0.204	0.807	0.677
32	VDEM Gender	0.697	0.215	0.820	0.682
33	WDI Education	0.673	0.217	0.758	0.617
34	WDI Resources	0.650	0.233	0.788	0.622

Table 8. Performance of models in the two year ahead prediction task

Rank	Model	Accuracy	Brier	AUROC	AUPR
1	Genetically optimized ensemble	0.839	0.122	0.928	0.845
2	Election Irregularities last election (short)	0.843	0.124	0.923	0.841
3	Election Irregularities (last election), VDEM exclusion, and WDI structural	0.841	0.124	0.906	0.808
4	Election Irregularities last election (long)	0.839	0.125	0.918	0.836
5	Election Characteristics last election (full)	0.836	0.126	0.917	0.828
6	Election Irregularities (last election), VDEM civil liberties, and WDI structural	0.841	0.126	0.906	0.812
7	Election Characteristics last election (full)	0.844	0.126	0.919	0.831
8	VDEM Mid level indicies, WDI structural, and DSP infrastructure	0.819	0.132	0.905	0.813
9	History of electoral violence (full)	0.811	0.133	0.912	0.821
10	VDEM Mid level indicies and WDI structural	0.826	0.134	0.903	0.804
11	VDEM full model	0.823	0.136	0.915	0.816
12	VDEM High level indivies and WDI structural	0.798	0.140	0.900	0.812
13	VDEM mid level indicies	0.807	0.140	0.912	0.830
14	DSP full model	0.802	0.140	0.912	0.800
15	DSP Social Media Climate, security	0.789	0.148	0.903	0.802
16	DSP Infra	0.805	0.148	0.890	0.785
17	Election Characteristics last election, structural	0.800	0.149	0.895	0.795
18	VDEM mid level indicies	0.799	0.149	0.897	0.794
19	DSP Disinformation and social media usage	0.807	0.151	0.897	0.800
20	DSP Disinformation and social climate and usage	0.799	0.152	0.896	0.795
21	DSP Social media climate	0.801	0.153	0.893	0.791
22	VDEM Political Exclusion Indicies	0.771	0.160	0.886	0.773
23	VDEM High level indicies	0.765	0.160	0.891	0.791
24	Full model (all features)	0.767	0.166	0.886	0.791
25	History of electoral violence (history only)	0.759	0.167	0.874	0.751
26	WDI Structural	0.760	0.173	0.856	0.708
27	VDEM Accountability Indicies	0.757	0.177	0.859	0.746
28	VDEM Neopatrimonialism	0.733	0.180	0.855	0.729
29	DSP Monitoring	0.741	0.181	0.865	0.727
30	VDEM Civil Liberties Indicies	0.729	0.192	0.838	0.717
31	WDI full model	0.735	0.203	0.814	0.676
32	VDEM Gender	0.701	0.210	0.819	0.689
33	WDI Education	0.661	0.220	0.749	0.606
34	WDI Resources	0.653	0.238	0.779	0.614