The Value of Local News Sources for Conflict Forecasting: Predicting Changes in Violence Intensity in Afghanistan Using Themes in the Media

Luuk Boekestein[1,†], **Jelke Bloem**[2] **and Jaap Kamps**[3]

[1] *Amsterdam University College*
[2] *University of Amsterdam*
[3] *University of Amsterdam*

Abstract

Contemporary efforts to predict conflict are still limited by interpretability issues and low accuracy in hard prediction cases. In the face of these limitations, the conflict predicting field has turned towards the use of big data such as large news databases, which has generally yielded promising results. Nevertheless, the blind application of news-based features for conflict prediction exposes forecasts to significant risks, and this work argues that more attention should be paid to careful data selection. To highlight this point, this study investigates the relative predictive performance of over 6 million local, regional, and global news articles scraped from the Global Database of Events, Language, and Tone (GDELT), by training 7 different models that aim to predict changes in conflict intensity in Afghanistan. The results show that local models significantly outperform regional and global models, especially on shorter-term predictions and in cases of more drastic change. This indicates that local media are better at capturing early signs of tensions than non-local media, and calls for a more diversified forecasting approach. Finally, this study argues that although big data approaches to conflict prediction might achieve higher accuracy, smaller-scale, interpretable, and conceptually justified approaches should be prioritized in order for the field to become a useful asset for preventing conflict.

Keywords: armed conflict, prediction, machine learning, bias, media

Corresponding author: Luuk Boekestein *Author email:* luuk.boekestein@gmail.com
Supervisor: Dr. Jelke Bloem *Supervisor email:* j.bloem@uva.nl
Reader: Dr. ir. Jaap Kamps *Reader email:* j.kamps@uva.nl
Submitted: May 29, 2024

Introduction

Being able to forecast the outbreak of violent conflict is one of the most crucial and fundamental tasks within peace research. Conflict prevention has become a central part of peacebuilding efforts by the United Nations (Muggah & Whitlock, 2022), and to this end, it has called for the development of early warning systems that could anticipate looming conflict. The potential of reliable forecasts to evaluate theory, compare policy responses, enable timely interventions and ultimately save lives, has motivated an ambitious field of research that aims to develop reliable forecasting methods (Bressan et al., 2019).

In recent years, researchers have turned towards the application of advanced machine learning methods and the use of big data approaches, and this has driven significant progress in the field. However, the field is still significantly limited in many ways, in particular due to the low interpretability of complex models (Ettensperger, 2019) as well as the low reliability of forecasting algorithms in hard prediction cases (Hegre, Metternich, et al., 2017; Hegre, Nygård, et al., 2017). Due to these challenges, the applicability of conflict prediction for peacebuilding on the ground is still limited. Several authors have tried to address these problems with creative methodological innovations, and one of the promising avenues is the incorporation of features from the media, which is posed to be able to help capture early warning signs and tensions in real-time (Chadefaux, 2014; Mueller & Rauh, 2022b). This has resulted in the use of large media-based event and sentiment databases for conflict prediction, which in some cases improved prediction accuracy, particularly in hard cases (Mueller & Rauh, 2022b).

However, the use of the media as a feature for conflict prediction is also accompanied by a significant risk. Forecasts could be based on misrepresentative data, biases, and fake news, especially since most media-based conflict prediction endeavors use mainly Western sources or mass-scraped news databases. Biased data can lead to biased forecasts, which is a seriously dangerous consequence due to the sensitive nature of violence forecasting applications. Therefore, several scholars have argued that the blind use of big databases under a "the more data the better" philosophy should be avoided, and the field should pay close attention to the limitations of big data (Cederman & Weidmann, 2017; Chadefaux, 2017; Mancini, 2013).

With the aim to contribute to this mission, this present work asks the following question: is local media better at predicting changes in violence than regional and global media? To investigate this, the differences in the forecasting accuracy of different kinds of news are evaluated by comparing the relative predictive power of models built with features from local, regional, and global news. The hypotheses are that (1) incorporating local media sources will lead to an increase in predictive performance compared to models without media information and (2) local media will be better at predicting conflict than regional and global media. This could furthermore underline the importance of a diverse forecasting framework, uncover biases in global conflict coverage, and highlight the value of local sources for conflict prediction.

Research Context

History of the field

Although the use of data methods for peace research has increased in the last decade, it is far from a new development. Since the early 1960s peace researchers have set out to systematically collect data on conflict, to enable the quantitative validation and development of theory, and to be able to predict conflict ahead of time (Small et al., 1997; Suzuki et al., 2002). It has long been argued that ultimately the ability to forecast peace and conflict is fundamental to the field (Hegre, Metternich, et al., 2017), and the ambitious mission was even dubbed by some early research as "the number one task of peace research" (J. David Singer et al., 1973).

Significant progress has been made since. Data on conflict has become increasingly more granular, due to important early contributions to systemic data collection (i.e. Azar (1980)), most prominently the KEDS[1] project of Schrodt et al. (1994) and the CAMEO[2] framework by Gerner et al. (2002). Fine-grained event data enabled the use of more advanced computational methods and this led to increasingly promising results (Schrodt, 1991, 2006; Schrodt et al., 2001; Subramanian & Stoll, 2006). These codifying methods have since been further developed and resulted in large-scale event databases such as the Integrated Crisis Early Warning System (ICEWS) (O'Brien, 2010) and the Global Database of Events, Language, and Tone (GDELT)[3] (Leetaru & Schrodt, 2013).

Still, the real boost for the conflict forecasting field came with the realization that conflict prediction methods play a crucial role in theory development. Conventional p-value-based quantitative studies have been argued to be highly misleading (Ward et al., 2010), and usually focus on explaining historical events at the cost of a tendency to overfit theoretical models to past events (Chadefaux, 2017). Through the lens of theories with little predictive relevance, "the future tends to surprise us" (Bressan et al., 2019, p. 6), and therefore many scholars argue that to evaluate theories, explanatory and predictive power are both necessary (Gleditsch & Ward, 2013; Hegre, Metternich, et al., 2017; Schrodt, 2013; Ward et al., 2010). Schrodt (2013) even lists the development of theory in the absence of prediction as one of the 'Deadly Sins' of quantitative political analysis. This marked a major turn in the field and research increasingly focused on the use of out-of-sample methods, which spurred many methodological innovations. Researchers also started drawing methods and practices from the machine learning field (Colaresi & Mahmood, 2017), which resulted in significant accuracy improvements and has pushed the field into the mainstream of conflict research (Bressan et al., 2019).

Limitations

However, despite the recent successes and steady improvements in accuracy, the applicability of conflict prediction for early warning systems (EWS), interventions and humanitarian response is still limited. Although it is well-established that early peacebuilding interventions can have a significant conflict-reducing effect (Hegre et al., 2019), there is a lack of a clear link between forecasted violence and potential interventions on the ground (Caldwell, 2022). Research into EWS in Africa has shown that the data-driven EWS is often too detached from local reality to inform meaningful early action responses (Eze & Osei Baffour Frimpong, 2020). As a result, a report on behalf of OCHA[4] concluded that the "current state of the art is not sufficient for application in humanitarian response" (Caldwell, 2022, p.11). This disconnect between research and practice significantly limits the usefulness of violence forecasting for early intervention or policy advice (Hegre, Nygård, et al., 2017)

This applicability gap has two main causes. Firstly, the introduction of complex machine learning methods into the field was accompanied by a trade-off between explanation and prediction accuracy (Hegre, Metternich, et al., 2017). More complicated models are more accurate, but often impossible to interpret (Hegre et al., 2022), leading to what Ettensperger (2019) labels as the "black box problem". This is furthermore complicated by the fact that the best-performing models are often a combination of different methods, known as ensemble methods, which are even harder to disentangle (Vesco et al., 2022). However, understanding the rationale behind a prediction is of pivotal importance for applicable results, and therefore the increase in accuracy achieved by complex but ambiguous methods might not be worth the loss of interpretability (Ettensperger, 2019; Hegre, Metternich, et al., 2017).

Secondly, predicting sudden changes in conflict, especially its onset after a period of no violence, remains difficult. Predicting the escalation of conflict after a long period of peace or the de-escalation after a long period of violence is something even the best-performing models struggle with (Bazzi et al., 2021; Hegre et al., 2022; Vestby et al., 2022) and has been labeled in the field as the "hard problem of conflict prediction" (Mueller & Rauh, 2022a). This is partly due to the dominance of past conflict as a predictor for future conflict (Caldwell, 2022), which means it is significantly easier to predict future conflict in a country that is already experiencing conflict (Hegre, Nygård, et al., 2017). In response, some authors have considered the possibility there might be theoretical limitations to predicting new conflict (Malone, 2022) or even that the presence of uncertainty is itself a requirement for war (Chadefaux, 2017), as has been posed before by political scientists (Gartzke, 1999).

Big data solutions

In attempts to address these limitations, in particular the hard problem, researchers have looked for data that could contain more information on early warning signs for conflict. Researchers have used remotely sensed data (Levin et al., 2018; Racek et al., 2023), Twitter (Dowd et al., 2020), and large-scale event databases (Attinà et al., 2022), but of particular interest has been the systemic extraction of information from the news. The news, and the media in general, is often posed to function as a mirror of society (Cook, 2000), and thus it has the potential to provide signs of political tension even before actual conflict has occurred (Chadefaux, 2014). Furthermore, the news could capture indicators for looming conflict at a way higher temporal resolution and in near real-time, which means it could provide more useful indicators of conflict compared to traditional structural variables, like GDP measures or democracy scores (Cederman & Weidmann, 2017).

Research has shown that the news indeed shows early warning signs of conflict (Chadefaux, 2014), displays signals of early tensions (Hegre et al., 2022), and enables the successful classification of countries into peace and conflict (Liebovitch et al., 2023). Of particular importance is the work of Mueller and Rauh (2017, 2022a, 2022b), who showed that the incorporation of newspaper articles in forecasting algorithms achieves a modest increase in prediction accuracy on hard cases. Moreover, many suggest that indicators derived from the media could become more important in the future, coinciding with a general shift of public discourse to the virtual realm (Bazzi et al., 2021; Dowd et al., 2020; Sweijs et al., 2022).

A database of particular significance for this purpose is the aforementioned GDELT project (Leetaru & Schrodt, 2013), which systemically archives worldwide online news in over 60 languages, and codifies themes, sentiments, tone, locations, events, and networks of actors (Saz-Carranza et al., 2020). It is recognized as the most comprehensive event and sentiment database that exists (Raleigh et al., 2023), and since its launch in 2011, the big data of GDELT has been applied to many areas in the social sciences. In the context of conflict prediction, it has been used to forecast social upheaval (Galla & Burke, 2018; Halkia et al., 2020; Keertipati et al., 2014; Korkmaz et al., 2015; Qiao et al., 2017), political risk (Sun et al., 2021), conflict intensity (Levin et al., 2018), global peace index (Voukelatou et al., 2022) and even the future of international relations as a whole (Chen et al., 2020). The response to GDELT has been enthusiastic, and some even explicitly argue that forecasting methods using its data could be useful for early interventions by peacekeepers (Voukelatou et al., 2022), and humanitarian organizations like the UN or the Red Cross (Keertipati et al., 2014).

On the other hand, critics argue that peace researchers should be wary of the blind application of big-data methods without carefully considering their nature and limitations. Research into large autocoded event databases, such as GDELT, has uncovered many limitations of its data, such as noise (Halkia et al., 2020), fake news (Raleigh et al., 2023), missing or duplicate data (Saz-Carranza et al., 2020), and a bias for urban events (Weidmann, 2016). Furthermore, despite efforts to diversify the sourcing of GDELT, most data consists

[1] Kansas Event Data System
[2] Conflict and Mediation Event Observations
[3] See www.gdeltproject.org
[4] United Nations Office for the Coordination of Humanitarian Affairs

of English and Western-based media (Saz-Carranza et al., 2020). Finally, there is a lack of transparency behind important classifying decisions such as "What counts as conflict?" (Day et al., 2015). Due to these limitations, Raleigh et al. (2023) argue that the claims by Keertipati et al. (2014) that GDELT can be used for humanitarian organizations to make "important decisions" are short-sighted.

In general, various scholars warn that although the incorporation of media features into conflict prediction opens up many opportunities, it also opens the door to biases and misrepresentation issues in the data (Raleigh et al., 2023). It is well established in the field of media studies that the media is subject to many biases, especially so in their coverage of conflicts (Weidmann, 2016; Wolfsfeld, 2011). Conflict coverage often contains a geographical bias (Barron & Sharpe, 2008) cultural bias (Barranco & Wisler, 1999), fake news (Day et al., 2015; Sweijs et al., 2022), censorship (Baum & Zhukov, 2015) and a bias towards violent events (Day et al., 2015). During conflict in particular, the societal mirror of the media is heavily distorted (Cook, 2000; Wolfsfeld, 2011), and selecting a representative collection of stories is hard, even manually (Davenport & Ball, 2002). Especially when the bulk of news that is used comes from mostly Western and English sources, events might be covered with certain prejudices (Demarest & Langer, 2018), not covered correctly (Demarest & Langer, 2022), or be missed entirely (Raleigh et al., 2023).

Using the right data is important, especially in the field of conflict prediction, problems in the data can lead to biased, uninformed, or manipulated forecasts, which could have dangerous consequences. (Raleigh et al., 2023), but the limitations of GDELT are often overlooked (Raleigh & Kishi, 2019). Due to the predictive dominance of conflict history, for example, models are often biased toward forecasting more conflict (Hegre et al., 2022). These models could misinform policy decisions with an expectation of conflict, creating self-fulfilling prophecies that perpetuate existing conflict cycles (Raleigh et al., 2023). Therefore, using mass-scraped media sources for conflict prediction could be dangerous (Sweijs et al., 2022), but the limitations of these large databases are often overlooked (Raleigh & Kishi, 2019).

Research focus

This research aims to address one of these limitations by investigating the relative predictive value of local, regional, and global news as present in the GDELT dataset. Due to data availability constraints and the scope of the project, this study will focus on a case study of Afghanistan between the years 2017 and 2024. Afghanistan was selected specifically due to its rich conflict history and because in these years it experienced both a period of war and a period of relative peace, which makes it an interesting case for a prediction algorithm. The monthly fatalities in Afghanistan are displayed in Figure 1. In order to evaluate the relative predictive performance of local, regional, and global news, this study used the so-called "themes" in the GDELT database as features for prediction. Despite the issues with the GDELT database, the themes identified may still be a relatively reliable source for media focus and tone (Raleigh et al., 2023), which is what this study is interested in.

The theme information was split into three distinct datasets that contain features from local, regional, or global media sources. To evaluate their predictive performance relative to other methods, the performance baseline was a simplistic model based on a set of historical features. Furthermore, three models were evaluated that combined both historical features and features from the themes, resulting in a total of seven models: a baseline model, three themes models (for local, regional, and global themes), and three history + themes models (for local, regional and global themes). This comparative approach could underscore the importance of careful data selection, and counter the idea that more data always leads to better predictions.

Methods

Historical features

For the information on conflict history, data from the Armed Conflict Location & Event Data Project (ACLED) was used. ACLED collects all types of information about violent events around the world, and is recognized as a reliable and complete source of conflict statistics with high validity (Raleigh & Kishi, 2019; Raleigh et al., 2010). The baseline model consists of a simple set of features based on historical violence and was adapted from the baseline model used by Mueller and Rauh (2022b). The historical feature set was chosen above other commonly used feature sets, such as economic factors, development measurements, or political indicators because it has consistently proven to be the most consistent predictor of violence (Hegre, Nygård, et al., 2017). The baseline model was given a total of ten features, related to the current and recent levels of violence (month t=0 till month t-3), the number of months since a certain level of violence (months since a month with 50, 500, and 2000 fatalities), and the aggregated number of fatalities over the last 3, 6 and 12 months.

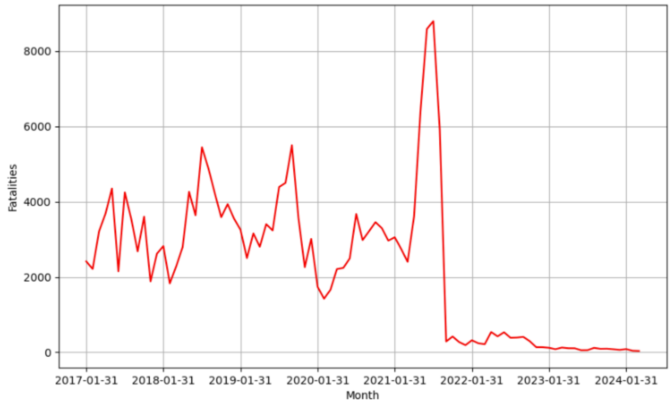

Figure 1. Monthly violence-related fatalities in Afghanistan between 2017 and 2024 *Note: Data derived from the Armed Conflict Location & Event Data (ACLED) Project, 2024. Retrieved from https://acleddata.com.*

Theme features

For information on themes present in the news, this study used the Global Knowledge Graph (GKG) from the GDELT project. The themes extracted[5] are based on a multitude of labels created by GDELT itself, and complemented with other taxonomies and labeling systems like those of the World Bank[6], CrisisLex[7], and the UN Global Pulse Project[8] (GDELT, 2015a, 2015b, 2016).

Using the GDELT DOC Application Programming Interface (API)[9], the total number of articles per day that mentioned Afghanistan were scraped, as well as the percentage of articles that were labeled with each theme out of the top 1000 most frequently used themes. The prediction was limited to a shorter period (January 2017 to March 2024) due to the limitations of GDELT's API. (GDELT, 2017). Consistent with the article selection criteria of Mueller and Rauh (2022b), news was considered to concern Afghanistan when it either mentions its name or the name of its capital, Kabul. Conveniently, GDELT automatically translated the search terms to all languages it supports, which means that global articles in over 65 languages were included

[5]The exact methods for GDELT's theme extraction is ambiguous and not-well documented, but seems to be based on the CAMEO framework by Gerner et al. (2002) and topic modelling techniques such as Latent Dirichlet Allocation (Saz-Carranza et al., 2020, p. 9). For a full list of themes present in the GDELT database, see https://blog.gdeltproject.org/new-august-2019-gkg-2-0-themes-lookup/

[6]See https://vocabulary.worldbank.org/taxonomy/1737.html

[7]See https://crisislex.org/

[8]See https://post2015.unglobalpulse.net/

[9]See https://blog.gdeltproject.org/gdelt-doc-2-0-api-debuts/

in the search. In total, information was scraped for 6.4 million articles published between January 2017 and March 2024, from a total of 162 countries. The degree of coverage of Afghanistan by all countries as present in the GDELT database is visualized in Figure 2.

The articles' theme information was then grouped into global (5.2 million articles), regional (941 thousand articles), and local news (201 thousand articles). An article was considered local when the source country was Afghanistan, and the news was classified as regional when the source country was Iran, Turkmenistan, Pakistan, Uzbekistan, Tajikistan, Azerbaijan, Kyrgyzstan, China, India, or Kazakhstan, due to their geographical proximity and substantial coverage of Afghanistan. The theme information was aggregated into local, regional, and global news, weighed by article count, and smoothed on a monthly basis.

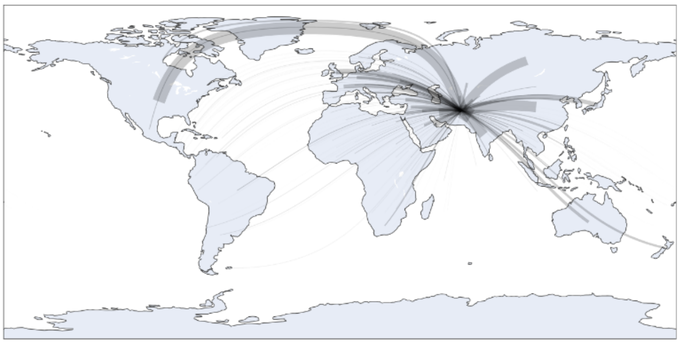

Figure 2. Coverage of Afghanistan by foreign countries, as present in GDELT

Figure 3 shows the total number of articles that were scraped for each of the news categories over time. A striking observation is that there is a significant spike in articles that concern Afghanistan from August 2021 to September 2021, which is when the US evacuation from Kabul took place (Stewart et al., 2021). It should be noted that this spike was only observable for global and regional news, as the article count of local news sources remained highly consistent throughout the whole period that articles were scraped.

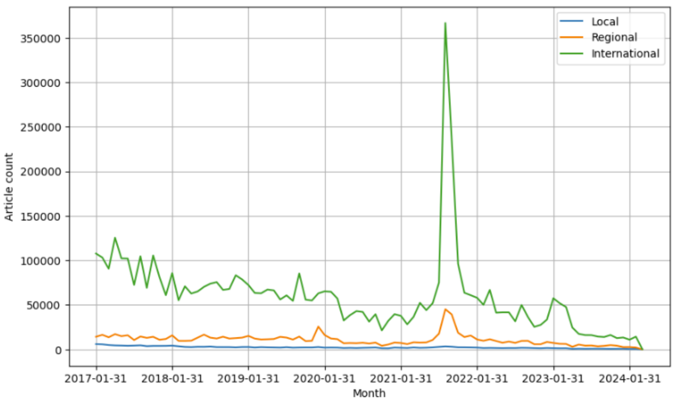

Figure 3. Monthly articles scraped

Finally, the 20 most relevant themes were selected for each model separately, by calculating the relative absolute correlation of each theme with the fatalities in month T+1. Due to the low sample size, it was found that this number of features led to the best performance, which is also generally the case (Mueller & Rauh, 2022a). The feature themes for each model are displayed in Table 1.

As can be seen in Table 1, most of the themes that had a high correlation with changes in conflict are clearly related to conflict, such as "Conventional War", "Insurgency", or "Conflict and Violence". Other themes seem more related to socioeconomic conditions, like "Economy", "Unemployment", or "Inflation". For most of the local themes, it is intuitively clear why the presence of this theme in the news would have something to do with conflict, but for the regional and global themes, this is not always obvious. Most prominently, the themes "World Languages Ukrainian" and "Ethnicity Ukrainian" in the global model, could have been picked up because the Russian invasion of Ukraine coincided with a period of lower conflict intensity in Afghanistan. This already indicates that the features picked up by the regional and global model may be of less conceptual relevance than the features picked up by the local model.

Prediction models

For the purpose of adhering to the standardized prediction frameworks established in the field set by Hegre et al. (2022), the goal was to predict changes in conflict intensity on the country-month (CM) level. Predicting the change in fatalities, as opposed to the exact number, is preferred because it often corresponds with the change in humanitarian impact, and is therefore more relevant for on-the-ground applications (Hegre et al., 2022). Change in conflict intensity is defined as the change in the log number of fatalities as a result of violence[10], and hence the prediction target O[11] becomes:

$$O_{s,i,t} = \Delta_s \ln(Y_{i,t+1}) = \ln(Y_{i,t+1}) - \ln(Y_{i,t-s}1)$$

where $Y_{i,t}$ is the aggregated number of fatalities in S months from now, and $Y_{i,t-s}$ is the aggregated number of fatalities in the current month. To predict one month into the future, the prediction target then becomes $\Delta_s \ln(Y_{i,t+1})$, which is the change in the log amount of fatalities plus one from month $t-1$ to month t (Hegre et al., 2022, p. 526). Due to the log transformation, the changes are predicted proportionally, and thus a shift from 10 to 20 fatalities is given the same severity level as a shift from 1000 to 2000 fatalities. Nevertheless, as argued by Mueller and Rauh (2022b), in practice, this is the same as first predicting the number of fatalities in the month $t+s$ and then calculating $O_{s,i,t}$ afterward, which is significantly easier than predicting the log change directly. Due to the one-month lag with which many conflict-prediction features become available, most models are only able to predict conflict at least 2 months ahead ($S = 2$), but given the real-time nature of the theme-related features (GDELT updates its database every 15 minutes), this research also made predictions for $S = 1$. Overall, violence was predicted up to six months into the future, corresponding to $S \in [1, 6]$.

For all the history models, the theme models, and the baseline model, random forest was used. Random forest is a popular machine-learning method, used for both classification and regression tasks. It ensembles a large number of randomly generated decision trees, and designs its cutoff points in such a way that the mean squared error (MSE) of the prediction is minimized. For the regression result, the average of all decision trees is computed. Even though more advanced machine learning methods have also been applied in the field, random forest remains among the most successful machine learning methods, and it is often preferred above more complex models due to its high interpretability. It has proven especially useful in the context of predicting violence with the news (Mueller & Rauh, 2017, 2022a, 2022b). Nevertheless, many other promising machine learning methods have recently been successful in the field, such as neural networks (Ettensperger, 2019), AutoML algorithms (D'Orazio & Yu, 2022), and ensemble methods (Ettensperger, 2021), but these were omitted in this study due to time and resource limitations.

The models were trained to minimize the MSE of the predicted log change in fatalities. Furthermore, parameters were optimized for each individual model using a grid search, which ran over the following parameters on a training split: number of trees (50,100,200), maximum depth of trees (none, 10, 20), minimum sample split (2,

[10] The definition of "fatalities" here follows ACLED's definition of "Reported Fatalities". See ACLED's codebook for more information: https://acleddata.com/acleddatanew/wp-content/uploads/dlm_uploads/2023/06/ACLED_Codebook_2023.pdf

[11] See Hegre et al. (2022) for detailed elaboration on the purpose of this prediction metric

Table 1. Themes used as features for local, regional and global models

Local News	Regional News	Global News
Conventional War	Insurgency	Insurgency
Political Violence and War	Rebels, Guerrillas, and Insurgents	Conventional War
Armed conflict	Women	Rebels, Guerrillas, and Insurgents
Insurgency	Inflation	Macroeconomic Vulnerability Debt
Security Services	Economy Inflation	Non-State Security Actors
Rebels, Guerrillas, and Insurgents	Conventional War	Economy Inflation
Conflict and Violence	Take Office	Inflation
Non-State Security Actors	Self-Identified Humanitarian Crisis	Food Security
Economy Historic	Reconciliation	Points of Interest Airport
Rebellion	Responses to Human Rights Abuses	Airports
Policy	Food Security	Self-Identified Humanitarian Crisis
Economy	Budget Deficit	Drainage
Religion Islamic	Tax Troops	Terror
Terror	Policy	Affordable Nutritious Food
Jobs	Affordable Nutritious Food	Political Violence and War
Deputy	Macroeconomic Vulnerability Debt	Security Services
Military	Unrest Checkpoint	Transport Infrastructure
Job Opportunities Employment	Non-State Security Actors	Drainage change
Unemployment	Gender Equality	World Languages Ukrainian
National Security	Checkpoint	Ethnicity Ukrainian

5, 10), and minimum sample at end leaf (1,2,4). For each prediction month t+s, the model was trained with all data up until the month t, with a base number of 12 months. In other words, the model with the least amount of data was trained on 12 months of data to predict changes in conflict intensity for month 13 to month 18, and the model with the most amount of data was trained to predict up to six months into the future beyond the dataset.

Results

Main results

Table 2 displays the mean squared error of all seven models. The optimal parameter set of each model can be found in Appendix A.

The first, and most obvious observation is that the MSE of all models increases substantially when the model has to predict for more months into the future. Nevertheless, some models suffer from this more than others, in particular the baseline model and the models history + local, history + global, and global. Interestingly, for some other models, the MSE seems to peak around the 3-4 month mark and then drop again slightly. The regional model, for example, achieves its second-lowest MSE around the 5-month mark.

Table 2. Overall MSE scores of prediction models

	Months ahead					
Model	1	2	3	4	5	6
Baseline (History)	0.540	1.134	1.683	2.303	2.839	2.914
History + Local	0.492	1.029	1.610	2.358	3.418	4.031
History + Regional	1.030	2.142	2.849	3.147	2.784	2.277
History + Global	1.032	1.627	1.911	2.871	3.338	4.150
Local	0.468	0.824	1.173	1.464	1.920	2.199
Regional	1.199	1.737	2.190	2.009	1.690	1.747
Global	1.398	1.608	1.731	2.216	2.765	3.326

Second, the results show that the history + local model slightly outperforms the baseline model. This is the case particularly for $S \leq 3$, as for the forecasts of more months in the future the baseline model performs better. Furthermore, both the baseline and the history + local model are outdone by the local model, which only uses theme information. The local model is in general the best-performing model for all $S < 5$ and has the lowest average MSE over all forecasting windows. Interestingly, for $S \geq 5$, the regional model achieves the lowest MSE.

Third, the results show that the local model achieves better scores compared to regional and global models, regardless of whether history was incorporated. This is the case particularly for the shorter forecasting windows, for longer periods this distinction is less well-defined. Moreover, it seems that on average, the regional models outperform the global models, although this depends on the forecasting windows and is mostly the case for $S \geq 4$. In general, the models that use historical features appear to be worse at predicting over longer time periods than the models that solely use theme information.

Comparing the baseline with local models

In the figures below, the relative performance of local models is explored in more detail. In Figure 4, the fatalities are plotted against the predicted fatalities of the baseline model, the history + local model, and the local model. Of particular interest is the ability of the models to anticipate the huge spike in fatalities that occurred in June and July of 2021. The baseline model seems to struggle with this drastic change and is mostly catching up with the increase in fatalities after the fact. Here, the local models do significantly better and are able to match the spike in fatalities more accurately, although still with a one-month lag. Furthermore, as seen in Figure 4, the local models are better at adjusting to the change from a period of intense conflict to a period of relative peacefulness. The baseline model predicts an increase in fatalities in the years after 2021, while the local models correctly predict the amount of fatalities to remain relatively stable.

The same phenomenon is observed when comparing local models to both the regional and global models. The MSE of the three different theme-based models are compared over time in Figure 5, which shows that although the three models are quite close in terms of MSE for the period up to August 2021, both the regional and the global model struggle to adapt to the period of relative peace that comes after.

Discussion

Interpretation of results

The first finding is that the incorporation of local media features leads to a slight increase in prediction accuracy, which was the first hypothesis. Local media is mostly better at shorter-term prediction, which is not unsurprising given that media features are posed to help catch

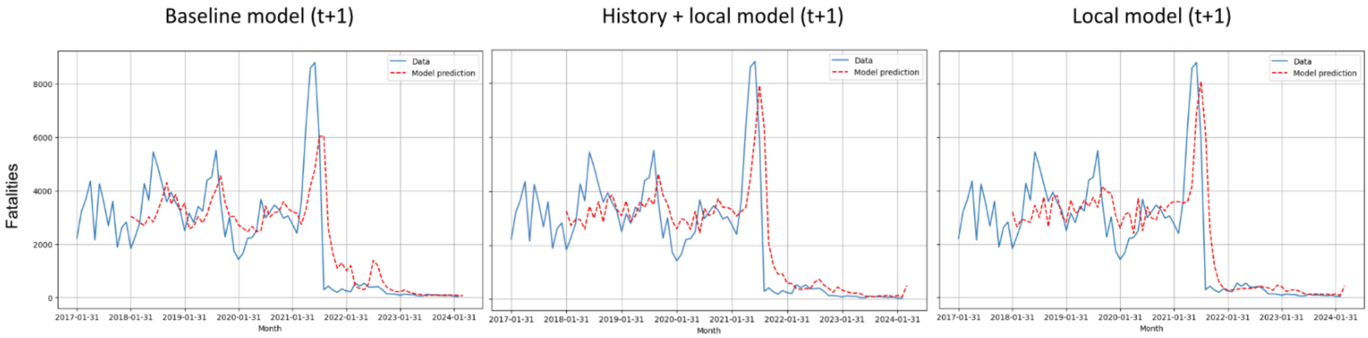

Figure 4. Comparison of local models with the baseline model

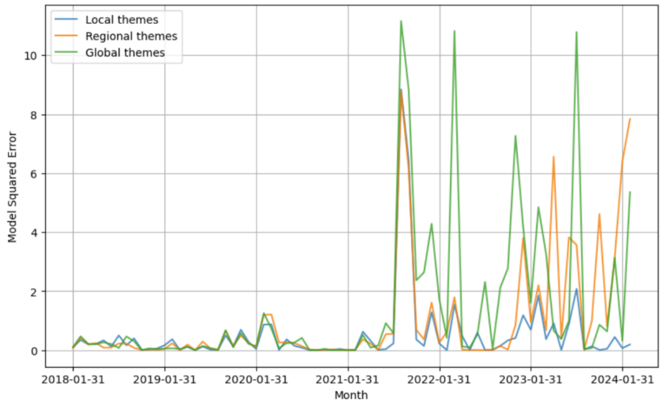

Figure 5. Where do local models outperform regional and global models?

signals on short-term tensions, not necessarily long-term developments (Chadefaux, 2014). The media features thus help improve the predictive accuracy of the baseline model, which is in line with earlier findings of Mueller and Rauh (2022a, 2022b). What is especially interesting about the results of this research is that the model that only uses theme features from the local media performs better than the local model that also uses history features. This could suggest that in hard prediction cases, historical features actually add noise to the model instead of contributing to it. Furthermore, in the period of relatively lower conflict intensity after August 2021, the local models provide more stable predictions, while the baseline struggles. This is likely because the training data mostly consisted of periods with much conflict, and thus the dominance of past conflict as a predictor for future conflict (Caldwell, 2022) could have influenced the model towards predicting more violence.

The second finding is that local media is significantly better at conflict forecasting than both regional and global media, confirming the second hypothesis. This indicates that locally sourced news is more valuable for deriving early signs of tensions than news from other countries. The local models are also noticeably better at anticipating the massive spike in fatalities in June 2021, constituting a slight improvement in accuracy for what is a hard prediction case and echoing earlier findings on the contribution of news data by Mueller and Rauh (2022a). Furthermore, local models particularly adjust better to drastic changes in conflict climate, something the baseline, regional models, and global models struggle with. As a result, the non-local models seem to suffer from a bias toward predicting violence in the period after 2021, which could be because international reporting of conflict often contains a violence bias in its coverage (Day et al., 2015). This is also illustrated by the fact that both regional and global media attention on Afghanistan spiked immensely in August 2021, coinciding with the spike in fatalities that occurred around the same time. The local media features provided more consistent coverage of the conflict, and this seems to have led to more accurate forecasts.

The third striking observation is that the model built on regional news features outperforms the model built on global news features in many cases. This suggests that the proximity of news sources matters and that the better performance could be explained by the higher geographical (Barron & Sharpe, 2008) and cultural (Barranco & Wisler, 1999) proximity of news sources in the region. A higher proximity may lead regional news sources to pay more attention to the situation in Afghanistan and therefore produce coverage that is more representative than global news sources. There have been attempts in the literature to quantify these measures into one, most prominently the notion of "relative proximity" as coined by Sheafer et al. (2014). The regional models still perform worse than the local models overall, but they seem to provide better predictions for longer forecasting windows, and therefore future models might benefit from a combination of both local and regional features.

Limitations of results

In general, the models in this research still suffer from the limitations most pertinent to all conflict prediction efforts. First of all, the accuracy of all the models drops immensely when predicting further into the future, and this remains a persevering symptom of forecasting efforts, as also communicated in recent state-of-the-art reviews (Hegre et al., 2022; Vesco et al., 2022). To a degree, this limitation is inherent to the problem by definition, but it nevertheless limits the degree to which predictions can be useful for real-life applications (Caldwell, 2022). Second, the hard problem remains hard, and even though the local models show slight improvements, all models failed to timely predict both the massive increase in fatalities following the US evacuation in 2021 and the massive decrease afterward. This aligns with prior studies on the hard problem (Bazzi et al., 2021; Hegre et al., 2022; Vestby et al., 2022), and confirms the early observation in the field that predicting conflict in a country that is already experiencing conflict is significantly easier (Hegre, Nygård, et al., 2017).

Furthermore, the comparison of the models with other studies in the field demonstrates that the performance of all seven models is relatively poor: state-of-the-art models in the field achieve significantly lower MSE scores, even those that employ a similar methodology. To contextualize the performance of the best model, Figure 6 compares the MSE scores of the local model of this study with the MSE scores of three high-performing models in the literature. All three of the comparison models were published as part of a forecasting competition organized by the Violence & Impacts Early-Warning System (VIEWS) project (Hegre et al., 2019), where participants were asked to predict changes in violence in Africa on a country-month level.

The first model is the news model by Mueller and Rauh (2022b), who employed a similar methodology as this study, using both measures of past violence and topics in the news to predict changes in violence. The second and third models are both ensemble models created by (Vesco et al., 2022), which consist of a weighted combination of 13 different models. The results are compared to both the

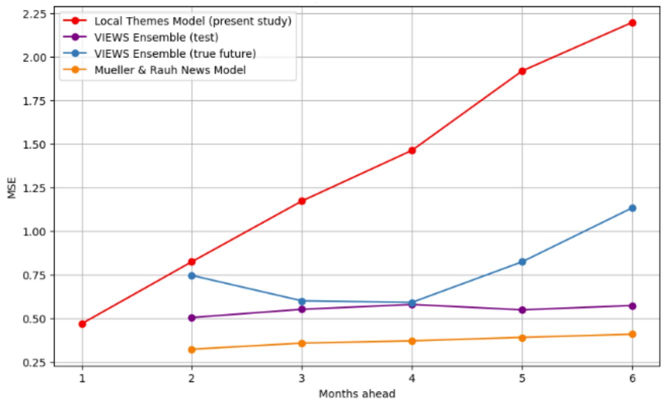

Figure 6. Comparison of the best model of the present study with state-of-the-art models

predictions of their model on a test-set, and to the predictions of their model for the true future[12][13].

Figure 6 clearly shows that the present model scores significantly worse than both the VIEWS models and the news model. For the first two forecasting windows $S \leq 2$, the local model still performs relatively well, but for the predictions further into the future the MSE of the local model increases linearly, while the MSE of the compared models remains relatively constant. This difference in prediction accuracy can partly be explained by the smaller data scope of this research. First, due to limitations of the GDELT API, the data was limited to a shorter time period compared to the models in the literature: the datasets for both VIEWS models and the news model range back to 1990, while the dataset for this research only ranges from January 2017 to March 2024. Second, most models in the literature use worldwide conflict features, but this research focused on a small case study of Afghanistan, further limiting the number of cases the model could draw from. This choice for a case study was motivated in large part by the desire to use local news, and the baseline was accordingly also constructed using only historical data from Afghanistan. Therefore, for a full comparison of the models presented here, either the dataset for the current methodology would have to be expanded, or the state-of-the-art would have to be recreated with similar, smaller datasets.

Finally, the model in this study was significantly limited by the quality of the data in GDELT's GKG. Other than the limited time window for which data was retrievable, many have identified problems with the nature of the data. Some of these, such as noise (Halkia et al., 2020), unclear conflict criteria (Day et al., 2015), and missing and duplicate data (Saz-Carranza et al., 2020), are mostly crucial for the GDELT event database and less so for the theme information derived from the GKG data used by the present models. However, other problems are more concerning, such as the presence of fake news (Raleigh et al., 2023), the disproportionate amount of Western and English media sources (Saz-Carranza et al., 2020), and the lack of robust documentation of the dataset.

Implications

Although the fact that news features add value to prediction efforts has been established before (Chadefaux, 2014; Liebovitch et al., 2023; Mueller & Rauh, 2017), the observation that local media contributes to a significantly greater extent to prediction accuracy introduces a new and important distinction. Crucially, filtering the large amounts of data scraped to only local data led to a better-performing model, which highlights that more data is not always better data. Media-related features will likely play a more pivotal role in the future (Bazzi et al., 2021; Dowd et al., 2020; Sweijs et al., 2022), but it is important that the field takes the value of carefully selected data seriously. Criteria for data selection should become a robust part of conflict prediction research, for a number of important reasons.

First of all, all data, and especially data that is derived from the news, is inherently sensitive to bias and subject to interference, by means of propaganda, misinformation, and censorship (Baum & Zhukov, 2015). This is illustrated by the many limitations of the GDELT dataset as discussed above, and although other sources of data may suffer from them less, to a degree all datasets of this nature face similar problems by definition (Chadefaux, 2017). Features from the news can indeed function as a mirror of society (Cook, 2000) and provide important early warning signs, but we have to carefully distinguish which mirrors are to be trusted and which are not. The combatting of bias, western dominance, and misinformation in data sources is thus paramount in order to use data for the prediction of future conflict, to avoid basing predictions on a distorted version of reality.

Second, the findings challenge the persistent notion in the field that more data is always better data or leads to more informed predictions. This resonates with existing calls by experts, who argue that the belief that "brute force" approaches using big data will lead to better forecasts is highly flawed in the domain of conflict studies (Cederman & Weidmann, 2017; Mancini, 2013). The overconfidence in machine learning methods and the sidelining of theoretical approaches is even dangerous, due to the complex nature of conflict environments (Mancini, 2013). Furthermore, Cederman and Weidmann (2017) argue that even if features from large-scale data-scraping operations have shown to contain signs of political tension, this does not necessarily mean that the resulting algorithms will be able to predict conflict with high accuracy.

These concerns are particularly relevant in light of the possibility that certain aspects of conflict, or even conflict as a whole, could be inherently unpredictable. The field has largely been under the assumption that since big data has been able to predict many other things that were once considered unpredictable, the current low accuracy in conflict prediction is attributable to methodological limitations (see i.e. Ward and Beger (2017)). However, the current limits for conflict predicting might not be due to flawed data or flawed models but could point to a fundamental theoretical limitation. As Chadefaux (2017) argues, perhaps conflict events are like clouds, highly unpredictable and irregular, or even like "Black Swans", and thus cannot be predicted at all (Taleb, 2010). The sudden surge in fatalities in Afghanistan in July 2021, for example, could simply have been so irregular that it was inherently impossible to predict. Another persevering hypothesis is that conflict might be in the "error term" of prediction (Gartzke, 1999), or that a high uncertainty about the future is itself a contributor to the onset of war (Malone, 2022). The conflict-predicting discipline has mostly overlooked the potential that conflict prediction might simply not be possible (Cederman & Weidmann, 2017), and seems driven by trust in data-driven progress. But the possibility of these limitations matters, because if the field is really trying to predict the unpredictable, then seemingly accurate predictions could lead to costly mistakes when they start to inform policy decisions (Chadefaux, 2017).

Therefore, it is important that data-driven approaches to the forecasting of conflict are theoretically informed, and that the field does not blindly fixate on naive accuracy measures such as MSE (Blair & Sambanis, 2020; Colaresi & Mahmood, 2017; Hegre et al., 2022). The field has to be careful not to mistake statistical straws in massive conflict databases for needles that represent some sort of conflict-predicting crystal ball. Big data can convey big errors (Taleb, 2013), and therefore Mancini (2013) argues that data-based approaches to conflict prediction should start on a small scale with local data before

[12] In the VIEWS competition, participants were asked to submit predictions for a test period, and for the true future at the time of submission (Vesco et al., 2022)

[13] Note that although the prediction task for the three compared models is similar enough to give some context to the results presented here, the task did differ substantially, and because it predicted changes in violence only for the African continent, no direct comparison with the predictions for Afghanistan of our model was possible.

drawing big conclusions on a larger scale. Researchers often opt to draw conflict features of many different countries and large timespans, and although this has been shown to increase the prediction accuracy (Hegre, Metternich, et al., 2017), this could also lead to self-fulfilling prophecies, or the repetition of one country's conflict dynamics in another (Chadefaux, 2021). Poorly designed conflict prediction algorithms are vulnerable to a bias toward predicting further conflict, as was shown in this research, and this could reinforce uninformed top-down decision-making structures, rather than enabling people-centered and bottom-up early warning systems (Muggah & Whitlock, 2022).

For a prediction to result in an effective intervention, it is not only important that a prediction is accurate, but also that it is understandable. To this end, simple models built on local data from the news might be more valuable, as they are easier to understand, and even though they might be less precise than more complex machine models, an increase in interpretability could outweigh slightly lower accuracy (Ettensperger, 2019, 2021). Current forecasting systems are often too remote and disconnected to be applied productively (Caldwell, 2022; Eze & Osei Baffour Frimpong, 2020; Hegre, Metternich, et al., 2017), but local news data could provide better information about the situation on the ground, and therefore help make forecasts more applicable.

The rationality of a conflict prediction is also important with respect to the warning-response gap that often prevents a forecast from resulting in effective policy intervention (Hegre et al., 2019; Musumba et al., 2021). If conflict prediction efforts want to make an impact on conflict prevention, it is crucial that early warnings are connected with suggested policy responses, and therefore forecasts have to be explainable and theoretically sound. Approaching a policymaker with the notion that "conflict will happen because Western media sources are writing about it", is unlikely to be persuasive or actionable. In this regard, using local news for a conflict prediction is inherently more appropriate for informing policy, and might be more conceptually desirable, regardless of the resulting accuracy. Of course, a prediction algorithm should not be deemed reliable just because it is based on local sources, but on the opposite end, an algorithm based on non-local sources warrants a significant dose of skepticism, no matter how accurate its prediction record.

Future works

This present work could be extended in the future by investigating the relative predictive value of news features in more detail. Firstly, future work should expand on the data collection, and deploy a more rigorous data-scraping methodology, for example by scraping news articles from different sources that might contain more local coverage, scraping over a larger time period, or predicting for multiple countries at the same time. Research could also focus on systemizing the distinction between regional and local news, for example by using relative proximity measures such as the one suggested by Sheafer et al. (2014). Furthermore, the performance of more complex machine learning methods that have proven successful in the field should be evaluated, in particular neural networks, AutoML algorithms, and ensemble methods. This could address some of the limitations of the existing study, increase the prediction accuracy, and enable a more direct comparison with the state-of-the-art in the field.

Finally, research should focus on explicitly interpreting the models' predictions, and evaluating to what extent they can be useful for informing policy decisions. There is some initial work with suggestions to improve the interpretability of more complex models, for example by Ward and Beger (2017) who provide some guidelines for the analysis of ensemble methods, and by Colaresi and Mahmood (2017) who designed a general framework for the critical analysis of machine learning research. This is also relevant with respect to the general mission in the field to incorporate quantitative forecasting methods into a larger framework with theory, qualitative methods, and human input (Chadefaux, 2017; Eze & Osei Baffour Frimpong, 2020; Hegre et al., 2022; Sweijs et al., 2022).

Conclusion

In closing, this research has highlighted the value of local news sources for predicting changes in conflict intensity. The comparison of the local, regional, and global models confirmed both hypotheses; (1) the incorporation of local media features leads to an increase in prediction accuracy compared to the baseline model, and (2) local media is significantly better at predicting conflict than both regional and global media. The models that use local media features are especially better at short-term prediction and at harder prediction cases. These findings suggest that local media sources could be more valuable for conflict prediction than regional and global media sources, and illustrate the value of careful data selection methods in the field. This presents a promising new avenue that could aid progress, especially because prediction frameworks using local news features could be better at capturing local political tensions, might suffer less from a bias towards conflict coverage, and are likely to be more conceptually sound and understandable.

Nevertheless, this present work has also shed light on a few significant limitations, and although the immense potential of conflict prediction is clear, whether we will ever be able to predict conflict with high accuracy is still up in the air. The current limitations could be on the methodological side or on the theoretical side, but either way, the prevailing assumption that big data methods can by themselves address a problem of such a complex nature is highly misinformed, and could have potentially disastrous consequences. This research suggests that it might be more fruitful to move in the other direction, and efforts should slow down and start on a smaller scale, to enable more careful consideration of diverse and local data collection methods. For the conflict prediction field to become more than just an interesting academic discussion, efforts should not aim merely for prediction accuracy, but also for high interpretability, conceptually justified methodology, and practical relevance for policy interventions. Only when predictions are coupled with meaningful tools for early intervention could these efforts hope to actually contribute to conflict prevention, and ultimately, save lives.

Replicability statement

The replication data and codes can be found under the authors Github page or by contacting the author directly, through the following links:
https://github.com/lboekestein/capstone
luuk.boekestein@gmail.com

REFERENCES

Attinà, F., Carammia, M., & Iacus, S. M. (2022). Forecasting change in conflict fatalities with dynamic elastic net. *International Interactions*, *48*(4), 649–677. https://doi.org/10.1080/03050629.2022.2090934

Azar, E. E. (1980). The Conflict and Peace Data Bank (COPDAB) Project. *Journal of Conflict Resolution*, *24*(1), 143–152. https://doi.org/10.1177/002200278002400106

Barranco, J., & Wisler, D. (1999). Validity and Systematicity of Newspaper Data in Event Analysis. *European Sociological Review*, *15*(3), 301–322. https://doi.org/10.1093/oxfordjournals.esr.a018265

Barron, P., & Sharpe, J. (2008). Local Conflict in Post-Suharto Indonesia: Understanding Variations in Violence Levels and Forms Through Local Newspapers. *Journal of East Asian Studies*, *8*(3), 395–423. https://doi.org/10.1017/S1598240800006482

Baum, M. A., & Zhukov, Y. M. (2015). Filtering revolution: Reporting bias in international newspaper coverage of the Libyan civil war. *Journal of Peace Research*, *52*(3), 384–400. https://doi.org/10.1177/0022343314554791

Bazzi, S., Blair, R. A., Blattman, C., Dube, O., Gudgeon, M., & Peck, R. (2021). The Promise and Pitfalls of Conflict Prediction: Evidence from Colombia and Indonesia. *The Review of Economics and Statistics*, *104*(4), 1–45. https://doi.org/10.1162/rest_a_01016

Blair, R. A., & Sambanis, N. (2020). Forecasting Civil Wars: Theory and Structure in an Age of "Big Data" and Machine Learning. *Journal of Conflict Resolution*, *64*(10), 1885–1915. https://doi.org/10.1177/0022002720918923

Bressan, S., Håvard Mokleiv Nygård, & Seefeldt, D. (2019). Forecasting and foresight: Methods for anticipating governance breakdown and violent conflict. *EU-LISTCO*, *1*(2). https://doi.org/10.17169/refubium-26117

Caldwell, S. (2022). *Assessing the Technical Feasibility of Conflict Prediction for Anticipatory Action* (tech. rep.). OCHA Centra for Humanitarian data. Retrieved January 1, 2023, from https://centre.humdata.org/assessing-the-technical-feasibility-of-conflict-prediction-for-anticipatory-action/

Cederman, L.-E., & Weidmann, N. B. (2017). Predicting armed conflict: Time to adjust our expectations? *Science*, *355*(6324), 474–476. https://doi.org/10.1126/science.aal4483

Chadefaux, T. (2014). Early warning signals for war in the news. *Journal of Peace Research*, *51*(1), 5–18. https://doi.org/10.1177/0022343313507302

Chadefaux, T. (2017). Conflict forecasting and its limits (T. Kuhn, Ed.). *Data Science*, *1*(1-2), 7–17. https://doi.org/10.3233/DS-170002

Chadefaux, T. (2021). A shape-based approach to conflict forecasting. *International Interactions*, *48*(4), 633–648. https://doi.org/10.1080/03050629.2022.2009821

Chen, P., Jatowt, A., & Yoshikawa, M. (2020). Conflict or cooperation? predicting future tendency of international relations. *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, 923–930. https://doi.org/10.1145/3341105.3373929

Colaresi, M., & Mahmood, Z. (2017). Do the robot: Lessons from machine learning to improve conflict forecasting. *Journal of Peace Research*, *54*(2), 193–214. https://doi.org/10.1177/0022343316682065

Cook, T. E. (2000). Governing with the News: The News Media as a Political Institution. *American Political Science Review*, *94*(3), 718–719. https://doi.org/10.2307/2585858

Davenport, C., & Ball, P. (2002). Views to a Kill: Exploring the Implications of Source Selection in the Case of Guatemalan State Terror, 1977-1995. *Journal of Conflict Resolution*, *46*(3), 427–450. https://doi.org/10.1177/0022002702046003005

Day, J., Pinckney, J., & Chenoweth, E. (2015). Collecting data on nonviolent action: Lessons learned and ways forward. *Journal of Peace Research*, *52*(1), 129–133. https://doi.org/10.1177/0022343314533985

Demarest, L., & Langer, A. (2018). The study of violence and social unrest in Africa: A comparative analysis of three conflict event datasets. *African Affairs*, *117*(467), 310–325. https://doi.org/10.1093/afraf/ady003

Demarest, L., & Langer, A. (2022). How Events Enter (or Not) Data Sets: The Pitfalls and Guidelines of Using Newspapers in the Study of Conflict. *Sociological Methods & Research*, *51*(2), 632–666. https://doi.org/10.1177/0049124119882453

D'Orazio, V., & Yu, L. (2022). Forecasting conflict in Africa with automated machine learning systems. *International Interactions*, *48*(4), 714–738. https://doi.org/10.1080/03050629.2022.2017290

Dowd, C., Justino, P., Kishi, R., & Marchais, G. (2020). Comparing 'New' and 'Old' Media for Violence Monitoring and Crisis Response: Evidence from Kenya. *Research & Politics*, *7*(3), 2053168020937592. https://doi.org/10.1177/2053168020937592

Ettensperger, F. (2019). Comparing supervised learning algorithms and artificial neural networks for conflict prediction: Performance and applicability of deep learning in the field. *Quality & Quantity*, *54*(2), 567–601. https://doi.org/10.1007/s11135-019-00882-w

Ettensperger, F. (2021). Forecasting conflict using a diverse machine-learning ensemble: Ensemble averaging with multiple tree-based algorithms and variance promoting data configurations. *International Interactions*, *48*(4), 1–23. https://doi.org/10.1080/03050629.2022.1993209

Eze, C. B., & Osei Baffour Frimpong. (2020). Contributions of Early Warning to the African Peace and Security Architecture: The Experience of the West Africa Network for Peacebuilding (WANEP). *Springer eBooks*, 181–194. https://doi.org/10.1007/978-3-030-46636-7_11

Galla, D., & Burke, J. (2018). Predicting Social Unrest Using GDELT. In P. Perner (Ed.), *Machine Learning and Data Mining in Pattern Recognition* (pp. 103–116). Springer International Publishing. https://doi.org/10.1007/978-3-319-96133-0_8

Gartzke, E. (1999). War Is in the Error Term. *International Organization*, *53*(3), 567–587. https://doi.org/10.1162/002081899550995

GDELT. (2015a, July). CrisisLex Taxonomies Now Available in GKG – The GDELT Project. Retrieved May 7, 2024, from https://blog.gdeltproject.org/crisislex-taxonomies-now-available-in-gkg/

GDELT. (2015b, March). World Bank Group Topical Taxonomy Now in GKG – The GDELT Project. Retrieved May 7, 2024, from https://blog.gdeltproject.org/world-bank-group-topical-taxonomy-now-in-gkg/

GDELT. (2016, January). UN My World Survey Topics Added to GKG 2.0 – The GDELT Project. Retrieved May 7, 2024, from https://blog.gdeltproject.org/un-my-world-survey-topics-added-to-gkg-2-0/

GDELT. (2017, June). GDELT DOC 2.0 API Debuts! – The GDELT Project. Retrieved May 7, 2024, from https://blog.gdeltproject.org/gdelt-doc-2-0-api-debuts/

Gerner, D. J., Rajaa Abu-Jabr, Schrodt, P. A., & Yilmaz, Ö. (2002). Conflict and Mediation Event Observations (CAMEO): A New Event Data Framework for the Analysis of Foreign Policy Interactions.

Gleditsch, K. S., & Ward, M. D. (2013). Forecasting is difficult, especially about the future. *Journal of Peace Research*, *50*(1), 17–31. https://doi.org/10.1177/0022343312449033

Halkia, M., Ferri, S., Papazoglou, M., Van Damme, M.-S., & Thomakos, D. (2020, May). Conflict Event Modelling: Research Experiment and Event Data Limitations. In A. Hürriyetoğlu, E. Yörük, V. Zavarella, & H. Tanev (Eds.), *Proceedings of the Workshop on Automated Extraction of Socio-political Events from News 2020* (pp. 42–48). European Language Resources Association (ELRA). Retrieved March 10, 2024, from https://aclanthology.org/2020.aespen-1.8

Hegre, H., Hultman, L., & Nygård, H. M. (2019). Evaluating the Conflict-Reducing Effect of UN Peacekeeping Operations. *The Journal of Politics*, *81*(1), 215–232. https://doi.org/10.1086/700203

Hegre, H., Metternich, N. W., Nygård, H. M., & Wucherpfennig, J. (2017). Introduction: Forecasting in peace research (H. Hegre, N. W. Metternich, H. M. Nygård, & J. Wucherpfennig, Eds.). *Journal of Peace Research*, *54*(2), 113–124. https://doi.org/10.1177/0022343317691330

Hegre, H., Nygård, H. M., Ræder, R. F., Metternich, N. W., & Wucherpfennig, J. (2017). Evaluating the scope and intensity of the conflict trap. *Journal of Peace Research*, *54*(2), 243–261. https://doi.org/10.1177/0022343316684917

Hegre, H., Vesco, P., & Colaresi, M. (2022). Lessons From an Escalation Prediction Competition. *International Interactions*, *48*(4), 1–34. https://doi.org/10.1080/03050629.2022.2070745

J. David Singer, Lear, J. B., Macaulay, D., & Meredith Reid Sarkees. (1973). The peace researcher and foreign policy prediction. *Papers of the Peace Science Society (International)*, *21*(1). https://doi.org/10.4324/9780203128398-18

Keertipati, S., Savarimuthu, B. T. R., Purvis, M., & Purvis, M. (2014). Multi-level Analysis of Peace and Conflict Data in GDELT. *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*, 33–40. https://doi.org/10.1145/2689746.2689750

Korkmaz, G., Cadena, J., Kuhlman, C. J., Marathe, A., Vullikanti, A., & Ramakrishnan, N. (2015). Combining Heterogeneous Data Sources for Civil Unrest Forecasting. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, 258–265. https://doi.org/10.1145/2808797.2808847

Leetaru, K., & Schrodt, P. (2013). *GDELT: Global Data on Events, Location and Tone, 1979-2012.* (tech. rep.). https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=20e2a85e4d2e8217d4693663e07f11de4dcd5a32

Levin, N., Ali, S., & Crandall, D. (2018). Utilizing remote sensing and big data to quantify conflict intensity: The Arab Spring as a case study. *Applied Geography*, *94*, 1–17. https://doi.org/10.1016/j.apgeog.2018.03.001

Liebovitch, L. S., Powers, W. J., Shi, L., Chen-Carrel, A., Loustaunau, P., & Coleman, P. T. (2023). Word differences in news media of lower and higher peace countries revealed by natural language processing and machine learning. *PLOS ONE*, *18*(11), e0292604–e0292604. https://doi.org/10.1371/journal.pone.0292604

Malone, I. (2022). Recurrent neural networks for conflict forecasting. *International Interactions*, *48*(4), 614–632. https://doi.org/10.1080/03050629.2022.2016736

Mancini, F. (Ed.). (2013). New Technology and the Prevention of Violence and Conflict. *Stability: International Journal of Security & Development*, *2*(3), 55. https://doi.org/10.5334/sta.cp

Mueller, H., & Rauh, C. (2017). Reading Between the Lines: Prediction of Political Violence Using Newspaper Text. *American Political Science Review*, *112*(2), 358–375. https://doi.org/10.1017/s0003055417000570

Mueller, H., & Rauh, C. (2022a). The Hard Problem of Prediction for Conflict Prevention. *Journal of the European Economic Association*, *20*(6), 2440–2467. https://doi.org/10.1093/jeea/jvac025

Mueller, H., & Rauh, C. (2022b). Using past violence and current news to predict changes in violence. *International Interactions*, *48*(4), 579–596. https://doi.org/10.1080/03050629.2022.2063853

Muggah, R., & Whitlock, M. (2022). Reflections on the Evolution of Conflict Early Warning. *Stability: International Journal of Security and Development*, *10*(1), 2. https://doi.org/10.5334/sta.857

Musumba, M., Fatema, N., & Kibriya, S. (2021). Prevention Is Better Than Cure: Machine Learning Approach to Conflict Prediction in Sub-Saharan Africa. *Sustainability*, *13*(13), 7366. https://doi.org/10.3390/su13137366

O'Brien, S. P. (2010). Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research. *International Studies Review*, *12*(1), 87–104. Retrieved March 11, 2024, from https://www.jstor.org/stable/40730711

Qiao, F., Li, P., Zhang, X., Ding, Z., Cheng, J., & Wang, H. (2017). Predicting Social Unrest Events with Hidden Markov Models Using GDELT. *Discrete Dynamics in Nature and Society*, *2017*, e8180272. https://doi.org/10.1155/2017/8180272

Racek, D., Thurner, P. W., Davidson, B. I., Xiao Xiang Zhu, & Göran Kauermann. (2023). Conflict forecasting using remote sensing data: An application to the Syrian civil war. *International Journal of Forecasting*. https://doi.org/10.1016/j.ijforecast.2023.04.001

Raleigh, C., & Kishi, D. R. (2019). *Similarities and Differences Across Conflict Datasets* (tech. rep.). ACLED. https://www.acleddata.com/wp-content/uploads/2019/09/ACLED-Comparison_8.2019.pdf

Raleigh, C., Kishi, R., & Linke, A. (2023). Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices. *Humanities and Social Sciences Communications*, *10*(1), 1–17. https://doi.org/10.1057/s41599-023-01559-4

Raleigh, C., Linke, R., Hegre, H., & Karlsen, J. (2010). Introducing ACLED: An Armed Conflict Location and Event Dataset. *Journal of Peace Research*, *47*(5), 651–660. https://doi.org/10.1177/0022343310378914

Saz-Carranza, A., Maturana, P., & Quer, X. (2020). *The Empirical Use of GDELT Big Data in Academic Research* (tech. rep. No. 822654). GLOBE. Retrieved March 9, 2024, from https://www.globe-project.eu/the-empirical-use-of-gdelt-big-data-in-academic-research_13809.pdf

Schrodt, P. A. (1991). Prediction of Interstate Conflict Outcomes Using a Neural Network. *Social Science Computer Review*, *9*(3), 359–380. https://doi.org/10.1177/089443939100900302

Schrodt, P. A. (2006). Forecasting Conflict in the Balkans using Hidden Markov Models. *Programming for Peace*, *1*(1), 161–184. https://doi.org/10.1007/1-4020-4390-2_8

Schrodt, P. A. (2013). Seven deadly sins of contemporary quantitative political analysis. *Journal of Peace Research*, *51*(2), 287–300. https://doi.org/10.1177/0022343313499597

Schrodt, P. A., Davis, S. N., & Weddle, J. L. (1994). Political Science: KEDS—A Program for the Machine Coding of Event Data. *Social Science Computer Review*, *12*(4), 561–587. https://doi.org/10.1177/089443939401200408

Schrodt, P. A., Gerner, D. J., Abu-Jabr, R., & Yilmaz, Ö. (2001). Analyzing the dynamics of international mediation processes in the Middle East and Balkans. Retrieved January 1, 2023, from https://www.semanticscholar.org/paper/Analyzing-the-dynamics-of-international-mediation-Schrodt-Gerner/03f8efbbef0c7015a09a42d728fcf2621c4f9013

Sheafer, T., Shenhav, S. R., Takens, J., & van Atteveldt, W. (2014). Relative Political and Value Proximity in Mediated Public Diplomacy: The Effect of State-Level Homophily on International Frame Building. *Political Communication*, *31*(1), 149–167. https://doi.org/10.1080/10584609.2013.799107

Small, M., J David Singer, & Bennett, R. (1997). *Resort to arms : International and civil wars, 1816-1980*. Umi, Stampa.

Stewart, P., Ali, I., Stewart, P., & Ali, I. (2021). Last U.S. troops depart Afghanistan after massive airlift ending America's longest war. *Reuters*. Retrieved May 9, 2024, from https://www.reuters.com/world/last-us-forces-leave-afghanistan-after-nearly-20-years-2021-08-30/

Subramanian, D., & Stoll, R. J. (2006). Events, Patterns, and Analysis Forecasting International Conflict in the Twenty-First Century. *Kluwer Academic Publishers eBooks*, 145–160. https://doi.org/10.1007/1-4020-4390-2_7

Sun, X., Gao, J., Liu, B., & Wang, Z. (2021). Big Data-Based Assessment of Political Risk along the Belt and Road. *Sustainability*, *13*(7), 3935. https://doi.org/10.3390/su13073935

Suzuki, S., Krause, V., & Singer, J. D. (2002). The Correlates of War Project: A Bibliographic History of the Scientific Study of War and Peace, 1964-2000. *Conflict Management and Peace Science*, *19*(2), 69–107. https://doi.org/10.1177/073889420201900204

Sweijs, T., Teer, J., van Hoeve, S., Berti, G., van Schie, J., van der Weerd, J., & Michaelis, J. (2022). Practices, Principles and Promises of Conflict Early Warning Systems. Retrieved January 1, 2023, from https://www.jstor.org/stable/resrep39641

Taleb, N. N. (2013). Beware the Big Errors of 'Big Data'. *Wired*. Retrieved May 27, 2024, from https://www.wired.com/2013/02/big-data-means-big-errors-people/

Taleb, N. N. (2010). *The Black Swan: The Impact of the Highly Improbable* (2nd ed., Random trade pbk. ed). Random House Trade Paperbacks.

Vesco, P., Hegre, H., Colaresi, M., Jansen, R. B., Lo, A., Reisch, G., & Weidmann, N. B. (2022). United They Stand: Findings from an Escalation Prediction Competition. *International Interactions*, *48*(4), 1–37. https://doi.org/10.1080/03050629.2022.2029856

Vestby, J., Brandsch, J., Larsen, V. B., Landsverk, P., & Tollefsen, A. F. (2022). Predicting (de-)escalation of sub-national violence using gradient boosting: Does it work? *International Interactions*, *48*(4), 841–859. https://doi.org/10.1080/03050629.2022.2021198

Voukelatou, V., Miliou, I., Giannotti, F., & Pappalardo, L. (2022). Understanding peace through the world news. *EPJ Data Science*, *11*(1), 1–30. https://doi.org/10.1140/epjds/s13688-022-00315-z

Ward, M. D., & Beger, A. (2017). Lessons from near real-time forecasting of irregular leadership changes. *Journal of Peace Research*, *54*(2), 141–156. https://doi.org/10.1177/0022343316680858

Ward, M. D., Greenhill, B. D., & Bakke, K. M. (2010). The perils of policy by p-value: Predicting civil conflicts. *Journal of Peace Research*, *47*(4), 363–375. https://doi.org/10.1177/0022343309356491

Weidmann, N. B. (2016). A Closer Look at Reporting Bias in Conflict Event Data. *American Journal of Political Science*, *60*(1), 206–218. https://doi.org/10.1111/ajps.12196

Wolfsfeld, G. (2011). *Making sense of media and politics : Five principles in political communication*. Routledge.

APPENDIX

A. Optimal grid search parameters

Model	n_estimators	max_depth	min_samples_split	min_samples_leaf
Baseline (History)	50	None	2	2
History + Local	50	None	2	1
History + Regional	100	None	2	1
History + Global	50	10	2	1
Local	200	10	2	1
Regional	200	10	2	1
Global	50	None	2	1