Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/20756
DOI URL: http://dx.doi.org/10.21474/IJAR01/20756

RESEARCH ARTICLE

PREDICTION OF INTERSTATE ARMED CONFLICTS IN WEST AFRICA USING ARTIFICIAL INTELLIGENCE TECHNIQUES.

Maurice Comlan[1], Horace Agossadou[2], Sardou Agbangla[1] andPatrick Sotindjo[3]

1. National School of Applied Economics and Management , University of Abomey-Calavi, 03 BP 1079, Cotonou, Benin
2. African School of Economics, 02 BP 372, Abomey-Calavi, Benin
3. National University of Sciences, Technologies, Engineering and Mathematics, BP 486, Abomey, Benin

Manuscript Info

Manuscript History
Received: 15 Febuary 2025
Final Accepted: 18 March 2025
Published: April 2025

Abstract

In the paper, we tested three machine learning models to predict the occurrence of inter-state armed conflicts in West Africa. The data used was collected from various sources over the period from 1981 to 2018. The modeling was done using Python. Among the three developed models, it turned out that the random forest model was the most suitable for this prediction. The modeling revealed that two major categories of variables are the most relevant predictors: whether or not countries share borders and the difference in democracy levels. Using the results of this prediction, we also identified the risk of conflicts between countries, such as Guinea vs. Sierra Leone, Burkina Faso vs. Benin, and Ivory Coast vs. Burkina Faso. This research deepens our understanding of state-to-state conflict dynamics in the West African region. However, it has limitations partly due to the lack of dynamic data availability.

Copy Right, IJAR, 2019,. All rights reserved.

Introduction.

The Sahel region, particularly Burkina Faso, Mali, and Niger, has experienced a significant escalation in organized political violence, primarily due to jihadist insurgencies. In 2021, Burkina Faso became the epicenter of the regional conflict, with organized political violence and deaths doubling compared to the previous year. This surge in violence has been largely attributed to Jama'at Nusrat al-Islam wal-Muslimin (JNIM), affiliated with Al-Qaeda, which has intensified its activities in several regions of the country [1]. The Niger also recorded a historic year of conflict in 2021, with the highest number of civilian deaths since ACLED (Armed Conflict Location & Event Data) data coverage began. The majority of these deaths were caused by the Islamic State's West Africa Province's Greater Sahara faction (ISWAP-GS), responsible for nearly 80% of all civilian deaths of the country[1]. West Africa is characterized by severe transnational vulnerabilities such as religious divisions and ethnic conflicts, which have exacerbated tensions and contributed to democratic backsliding. In Nigeria, for example, the ISIS West Africa Province has extended its reach beyond its traditional base, reflecting a broader trend of increasing jihadist activity across the region [2]. The ongoing military juntas in Burkina Faso, Mali, and Niger have significantly influenced the region's security policies. These regimes have adopted militarized counter-insurgency approaches, which have often led to increased violence and instability. The involvement of international forces, including French troops and the Wagner Group, has also shaped the security landscape, with mixed impacts on regional stability [3]. The central

Corresponding Author:-Maurice Comlan
Address:-National School of Applied Economics and Management , University of Abomey-Calavi, 03 BP 1079, Cotonou, Benin.

Sahel states continue to face a complex array of challenges, including military coups, escalating jihadist insurgencies, and widespread violence.
The application of machine learning in conflict prediction effectively serves as an early warning sys- tem, allowing proactive measures before violence escalates. Using computer programs, the international com- munity can anticipate and act to prevent potential conflicts. This predictive capability is crucial in regions like West Africa, where patterns of political instability and violence can be identified and addressed before caus- ing widespread harm [4, 5]. Moreover, identifying key conflict predictors provides a clear set of political and economic levers that could potentially prevent future unrest. Variables that poorly predict conflicts, such as rainfall or income inequality, are less likely to be immediate causes of interstate conflict, guiding the direction of preventive measures. Several regional organizations, such as the Economic Community of West African States (ECOWAS), have developed data collection mechanisms to help political authorities make informed de- cisions. For example, in June 2018, ECOWAS signed a Memorandum of Understanding with the West African Economic and Monetary Union (UEMOA) to strengthen crisis and conflict prevention in the West African region [6]. In terms of cross-sectional analysis, challenges remain considerable, as evidenced by the conflict systems in the Sahel or the Lake Chad Basin, which require regular updates of tools. This mismatch between analysis indicators and regional instability factors has long been a weakness of early warning systems [7]. De- spite recent developments toward broader democratic spaces, the region remains prone to tensions and crises. Anticipating conflicts can reduce the number of civilian and military casualties. Prevention also avoids human losses and suffering associated with armed conflicts. This prevention will also protect resources, trade routes, and investments.

The escalation of interstate armed conflicts in West Africa presents a significant challenge to regional stability, economic growth, and human security. Existing early-warning systems, such as those developed by organizations like ECOWAS, rely on traditional data aggregation and analysis methods, which often fail to capture the complexity and dynamics of conflict triggers. While prior research highlights factors like economic disparities, political instability, and territorial disputes as contributors to conflicts, these approaches lack pre- dictive precision and adaptability. Major limitations include insufficient integration of dynamic data, reliance on static indicators, and challenges in forecasting low-frequency but high-impact events. Through this research, we aim to leverage machine learning models to enhance the predictive accuracy of conflict occurrence, identify key predictors such as democracy level differences and border-sharing, and propose data-driven strategies for early intervention and conflict prevention. By addressing these gaps, we seek to provide policymakers with robust tools for proactive and informed decision-making.

This study investigates the use of machine learning models to predict interstate armed conflicts in West Africa. While previous studies have explored the impact of political instability, economic disparities, and territorial disputes on conflict dynamics, they did not explicitly address the integration of these factors into predictive models capable of generating actionable insights. Additionally, earlier research often lacked dynamic data and failed to capture low-frequency but high-impact events, leaving a critical gap in effective conflict forecasting and prevention strategies. The overall objective of this study was to use machine learning techniques to predict the occurrence of interstate armed conflicts in West Africa. Specifically, it aims to: i) identify models capable of better predicting the occurrence of interstate armed conflicts; ii)identify factors that have historically played a crucial role in predicting armed conflicts and iii) present some predictions of interstate armed conflicts in West Africa. To achieve these specific objectives, the following hypotheses were formulated: H1: The random forest model better predicts the occurrence of interstate armed conflicts; H2: The occurrence of armed conflicts between two states is due to the increase in Gross Domestic Product (GDP) ratio between countries sharing the same borders and H3: There are West African countries sharing the same borders with probabilities of armed conflict occurrence greater than 0.5 in the last 10 years. This article is organized into three sections. Section 1 presents the state of the art and related works on this topic. Section 2 presents the material and methodology used. Section 3 presents the results and a discussion.

Related Works.

Wars and armed conflicts are ubiquitous in current affairs and affect state security and human security on all continents. Hardly a day goes by without mentioning the outbreak, progression, or, more fortunately, the suspension or cessation of a war or conflict [8]. The propensity for wars between states, characterized by armed conflicts between two or more nations, stems from a complex interaction of factors such as territorial disputes, economic considerations, regime types, military strength, and geopolitical and democratic interests.
Despite recent developments marked by the relative expansion of democratic spaces, West Africa remains a region particularly affected by latent or violent conflict hotspots, notably related to the presence or activity of armed groups. This is generally a contrasted situation where tendencies towards entrenched conflicts coexist with dynamics of stabilization or crisis resolution. However, most countries in the region escape open conflict situations. Still, episodic tensions((protests against high living costs, pre-electoral or post-electoral crises) remind us of the risks of sliding into open crisis situations [7]. In the era of globalization, recognizing and understanding these causes is essential to inform policies and initiatives aimed at preventing conflicts. Among the various prediction methods and techniques, machine learning has proven effective. Machine learning is a rapidly evolving field increasingly adopted by various scientific disciplines.

Machine Learning (ML), a branch of artificial intelligence (AI), transforms traditional statistical meth- ods into predictive and interpretive models, particularly with the rise of Big Data. Machine learning algorithms, such as random forests [9] and logistic regression [10], offer significant advantages in analyzing large, complex datasets. For example, random forests, by aggregating multiple decision trees, overcome the weaknesses of in- dividual decision trees and improve predictive accuracy. Logistic regression, used for binary variables, is ideal for complex analyses in conflict studies. Decision trees, providing a clear visualization of decision rules, cap- ture non-linear relationships and interactions between predictors. These techniques enhance the prediction and analysis of interstate conflicts, helping to better understand underlying factors and develop effective preventive policies.

The use of artificial intelligence (AI) in armed conflict prevention is gaining increasing interest within the scientific community. Researchers are exploring various potential applications, such as predictive analysis of geopolitical tensions, early detection of conflict signs, and modeling de-escalation scenarios. AI systems can process vast amounts of data from diverse sources (social media, diplomatic reports, economic indicators) to identify patterns and trends likely to indicate imminent escalation. Some scientists are also working on algorithms capable of suggesting optimal mediation strategies based on the analysis of past conflicts. However, the community remains divided on the actual effectiveness of these tools and the ethical questions they raise, notably regarding potential biases and automated decision-making in contexts as sensitive as international peace and security. These works demonstrate the diversity of approaches and perspectives on using AI in armed conflict prevention, ranging from long-term prediction to practical conflict management. In the field of AI use for armed conflict prevention, many researchers make significant contributions. The project ViEWS [11] uses machine learning to predict conflicts up to 2050, considering climate scenarios. Some studies explore the legal and ethical implications of AI in UN peacekeeping operations [12]; the use of AI to identify early warning signs of genocides [13]; the potential of AI in conflict mediation [14]; the application of AI in tracking illegal arms flows, thereby contributing to conflict prevention [15]; and how AI could transform deterrence strategies [16]. Together, these works illustrate the richness and diversity of approaches in this rapidly expanding field, ranging from long-term prediction to practical conflict management, including ethical and strategic considerations.

Materials and Methods.

This section addresses the processes through which data were collected, prepared, and cleaned. Sub- sequently, we will provide a justification for the selection of three models, namely the random forest, logistic regression, and decision tree. Finally, we will describe the methodology used to evaluate the performance of these models.

Material

For this research, we used the Python programming language ecosystem with libraries such as Pandas, Numpy, Matplotlib, Patsy, Statsmodels, Skforecast, among others. Due to the importance of the analyses and the size of the data, the software suite was installed on a computer with the following technical specifications: Windows 10 Professional operating system, 64-bit, 8GB RAM, 2TB ROM, Intel® Core™ i5-1035G1 CPU @ 1.00GHz 1.19GHz processor. To predict an interstate conflict, it is essential to discern the underlying causes of conflicts to select relevant variables for constructing datasets. As highlighted in the literature review, economic, military, and democratic factors are often at the forefront of these causes [17]. Based on this understanding, our dataset incorporates key variables presented in Table 1. The database comprises a total of 392,135 observations from 1981 to 2018 and covers 156 out of the 191 countries worldwide. Regarding West Africa, the data encompass all 16 countries in the region, namely: Benin, Burkina Faso, Cape Verde, Ivory Coast, The Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Mauritania, Niger, Nigeria, Senegal, Sierra Leone, and Togo.

To predict interstate conflicts, it is essential to discern the underlying causes of conflicts in order to select relevant variables for building datasets. As we have seen in the literature review section, economic, military, and democratic factors are often at the forefront of these causes [17]. Based on this understanding, our dataset incorporates the key variables presented in Table 1. The period covered by the data is from 1981 to 2018. This period was chosen primarily due to the availability of data in the sources consulted. This temporal coverage encompasses a variety of events and geopolitical changes, making it suitable for accurately predicting interstate conflicts, given the dynamic nature of interstate relations. The military expenditure variable had 103 missing values. After data cleaning, we worked with a total of 392,033 observations, as indicated in the descriptive summary Table 1. The complete database is accessible online [18].

Table1. StatisticsofQuantitativeData

VARIABLES	MEAN	STD	MIN	25%	50%	75%	MAX
YEAR	2001.16	10.38	1981	1993	2000	2010	2018
CONFLICT	0.00	0.01	0.00	0.00	0.00	0.00	1.00
INSTABILITY_PO	0.08	0.28	0.00	0.00	0.00	0.00	2.00
GDP	15.98	107.24	0.00	0.12	0.73	4.40	7532.61
MILITARY_EX	13.14	2207.61	0.00	0.49	0.96	1.93	961947.81
DEMOCRACY_DIFF	7.26	5.87	0.00	2.00	6.00	12.00	20.00
SHARE_BORDER	0.02	0.15	0.00	0.00	0.00	0.00	1.00

Regarding the variable of interest in our research (CONFLICT), the statistics illustrate a low occur- rence of interstate armed conflicts worldwide, with only 46 cases recorded during the studied period. Among the 46 recorded global conflict cases, 13 occurred between countries sharing borders, compared to 33 cases for countries not sharing borders. We also observe that the level of democracy is generally lower in countries that have experienced the studied phenomenon. Political instability seems to be an influential factor in the occurrence of this event. The analysis of correlation and multicollinearity between our explanatory variables and the dependent variable confirmed the relevance of their inclusion in our models (Figure 1).

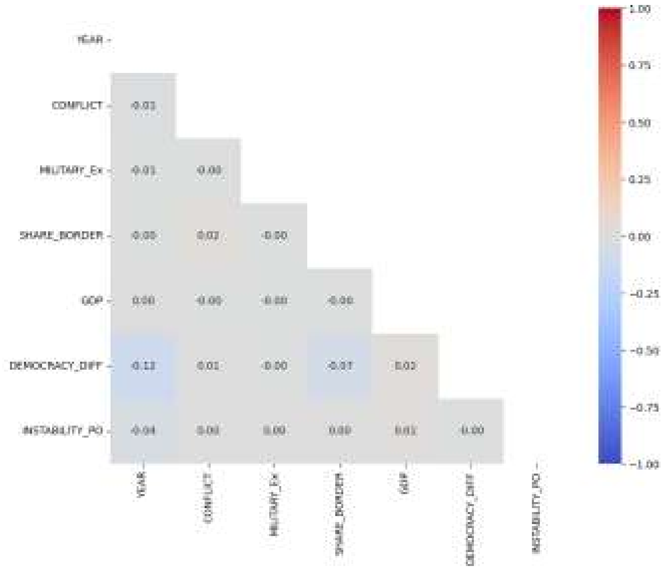

Figure 1. Pearson Correlation Matrix

It appears that the correlation between the variable of interest and the other variables is almost nil. There is a weak but negative correlation between sharing borders and the difference in the level of democ- racy between two countries. Regarding the multicollinearity test, the obtained VIF (Variance Inflation Factor) coefficients are all less than or equal to 1. Generally, a VIF greater than 5 or 10 may indicate strong multi- collinearity, but our values, being around 1, indicate excellent independence among each predictor. The data covers the period from 1981 to 2018. This period was chosen mainly due to the availability of data from the consulted sources. This period encompasses a variety of events and geopolitical changes, making it suitable for accurate prediction of interstate conflicts, given the dynamic nature of interstate relations.

We organized our data into dyads, which improves the predictive approach. According to this method, each country is paired with all others, allowing for a more effective comparison of the data, as highlighted by Ward [19]. During the initial exploration phase, we noticed that some countries had many missing values for the variable MILITARY Ex. Such missing data problems, if not resolved, can compromise the validity and reliability of statistical analyses [20]. The combination of their absence from the main event of interest (war) and the prevalence of missing values suggests that their inclusion could introduce noise rather than clarity, a concern noted in previous conflict prediction studies [21]. Furthermore, regarding data on military expenditures, GDP ratio, and the difference in the level of democracy, we adopted the data normalization method suggested by Singh in 2020.

This approach involves adjusting all variables to the same scale, often between [0,1], to prevent any variable from exerting a disproportionate influence on the model due to its amplitude. This step contributed to increasing the robustness of our models. Detailed preprocessing steps, scripts, and additional data visualizations are provided in the supplementary materials and referenced where relevant for replication.

Methodology and Hypothesis Testing

The analysis of the data we collected involves modeling using random forest, logistic regression, and decision tree methods to estimate the probability of the occurrence of an armed conflict between pairs of countries in West Africa. To do this, we relied on a variety of global socio-economic and geopolitical indicators, as regional data are insufficient. The explained variable, or labeled variable, is the ”CONFLICT” variable. Our different models include four (4) explanatory variables: INSTABILITY PO, GDP, MILITARY Ex, DEMOCRACY DIFF and SHARE BORDER.

Testing Hypothesis 1 : The verification of Hypothesis 1 involves evaluating the models. This will determine the model(s) best able to predict the occurrence of armed conflicts. There are multiple evaluation indicators, among which the most common are certainly ”precision,” the ROC curve, and the associated AUC indicator. While precision is often seen as intuitively relevant and the ROC/AUC score is considered a gold standard, these indicators have their own limitations, particularly in relation to the specific problem being addressed. Precision is no- tably unsuitable for imbalanced datasets. The ROC/AUC score is often criticized for overestimating model performance on highly imbalanced datasets [22]. In our case, ”Recall” and ”Precision” indicators, as well as the ”Precision-Recall Curve” and two other closely related ratio measures, proved relevant. The comparison metrics used are: Accuracy, Precision, Recall, F1 Score, and the confusion matrix.

Testing Hypothesis 2 : Determining the predictors of conflicts involves determining the importance of the features in our models. Feature importance refers to techniques that calculate a score for all input features for a given model. The scores represent the ”importance” of each feature. A higher score means that the specific feature will have a greater effect on the model used to predict a certain variable. In our study, we will retain, for each model deemed useful for conflict prediction, the predictors with the highest scores. If an assigned coefficient is a high number (negative or positive), it has a certain influence on the prediction. Conversely, if the coefficient is zero, it has no impact on the prediction.

Testing Hypothesis 3 : The analysis to verify this hypothesis involves examining probabilities predicted by the best-performing model. The process includes retrieving predicted probabilities from 2009 to 2018 and selecting dyads of West African countries sharing borders. For each dyad, the highest conflict probabilities are identified, duplicates are filtered, and the probabilities are ranked in descending order.

Results and Discussion.

The essence of our study lies in the modeling, evaluation, and forecasting steps we conducted. This section aims to reveal the results obtained from the application of the aforementioned modeling methods. We began our exploration with three modeling techniques. Each technique was rigorously tested based on a range of performance indicators. Finally, we presented a summary of the projections made by these models, specifically for the West African region.

Random Forest-Based Model

A random forest consists of multiple independent decision trees. This method incorporates two forms of randomness: first, each tree is developed from a random sample of the base data; second, a random subset of features is chosen to determine the best split at each node of the tree. This approach leverages the advantages of decision trees while reducing their propensity to overfit, thanks to the training of individual trees on different subsets of the data. For our data, we started by using the Gini index as the impurity measure, followed by a grid search to fine-tune the hyperparameters. Table 2 illustrates the performance obtained with this implementation of the model.

Table2.PerformanceMeasuresoftheRandomForestModel

Measures	Values
Accuracy	98%
Precision	100%
AUC	80%
Recall	83%
F1Score	89%

These indicators reveal that the random forest model performs exceptionally well on the test set, dis- playing excellent accuracy, remarkable precision, and a satisfactory balance between recall and discriminatory capacity (AUC), as indicated in Figure 2 (left). Indeed, the model correctly predicted 98% of the cases, effec- tively distinguishing between positive and negative instances of interstate conflicts. The ROC curve (shown in red in Figure 2, left) illustrates the trade-off between the true positive rate (sensitivity) and the false positive rate at different decision thresholds. In summary, it has been shown that our random forest model has a good capacity to differentiate between conflict and non-conflict classes.

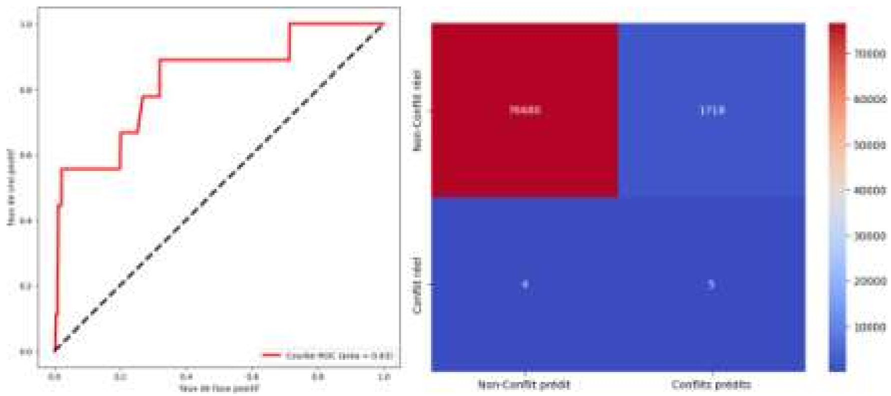

Figure 2. ROC Curve and Confusion Matrix of the Random Forest Model

Logistic Regression Model

The second model we implemented is logistic regression, which belongs to supervised learning meth- ods. Just like the previous model, we determined the metrics of the logistic regression model, which are detailed in Table 3. This model shows an accuracy and recall of 79%, despite class imbalance. This indicates that the model is capable of correctly predicting about eight out of ten armed conflicts. The AUC index confirms that this model has good discriminatory capability, with an 80% chance of correctly distinguishing between a posi- tive and a negative observation taken at random. This performance is corroborated by the ROC curve associated with this metric, illustrated in Figure 3 (left). Although the model is capable of correctly predicting 80% of the cases, the confusion matrix reveals a considerable number of false positives (67,142). This high rate may be a concern in some scenarios, as falsely predicting a conflict that does not exist could have serious consequences.

Table3.PerformanceMeasuresoftheLogisticRegressionModel

Measures	Values
Accuracy	79%
Precision	100%
AUC	80%
Recall	79%
F1Score	88%

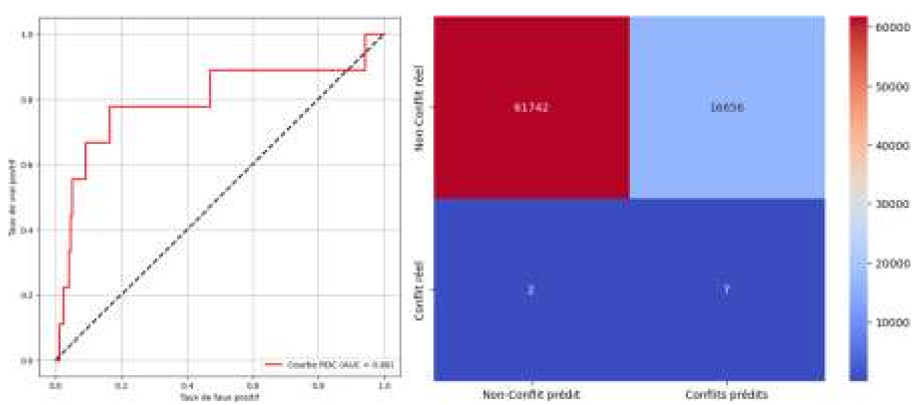

Figure 3. ROC Curve and Confusion Matrix of the Logistic Regression Model

Decision Tree Model

The aim of this method is to develop a model capable of predicting the value of a target variable based on simple decision rules inferred from the data features. A tree can be seen as a form of piecewise constant approximation. The performance of this model according to various metrics is presented in Table 4. Analysis of this table reveals that the model systematically predicts the occurrence of interstate armed conflicts (precision = 1). The high values of recall (0.95) and accuracy (0.95) suggest that the model detected almost all actual conflict cases in the test set.

Table4.PerformanceMeasuresoftheDecisionTreeModel

Measures	Values
Accuracy	95%
Precision	100%
AUC	53%
Recall	95%
F1Score	97%

The lowest measure, AUC, might reflect a class imbalance problem. This low AUC value suggests that the classifier may not be very effective in predicting interstate armed conflicts, as its results are close to those obtained by chance, as shown by the ROC curve (see Figure 4). The confusion matrix in this figure (right) provides a detailed account of the model's predictions, indicating that it correctly identified 1 out of 8 conflict cases and recognized a significant number of events as interstate armed conflicts. Although the number of false negatives is relatively low, in the field of conflict prediction, each undetected case can have serious consequences.

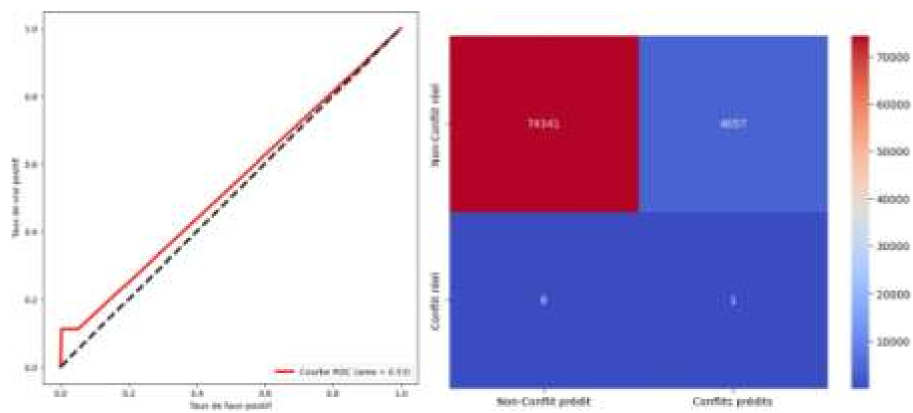

Figure 4. ROC Curve and Confusion Matrix of the Decision Tree Model

However, the decision tree model revealed, in decreasing order of importance, that the level of democ- racy, GDP ratio, and border-sharing between two countries are very significant indicators for predicting the emergence of interstate armed conflicts. Political instability and military expenditures have lesser relevance in this context. With the decision tree model, we created a tree (Figure 5) that helps us classify our observations. This visualization provides a clear graphical representation of the decision-making process, as mentioned ear- lier. The root node of the tree split based on the difference in democracy levels, with a threshold below 0.925, constituting the first level of decision. The choice of the root node is based on the feature that provides the greatest impurity reduction during the first split. In our study, the difference in democracy levels between countries allowed for the best initial separation.

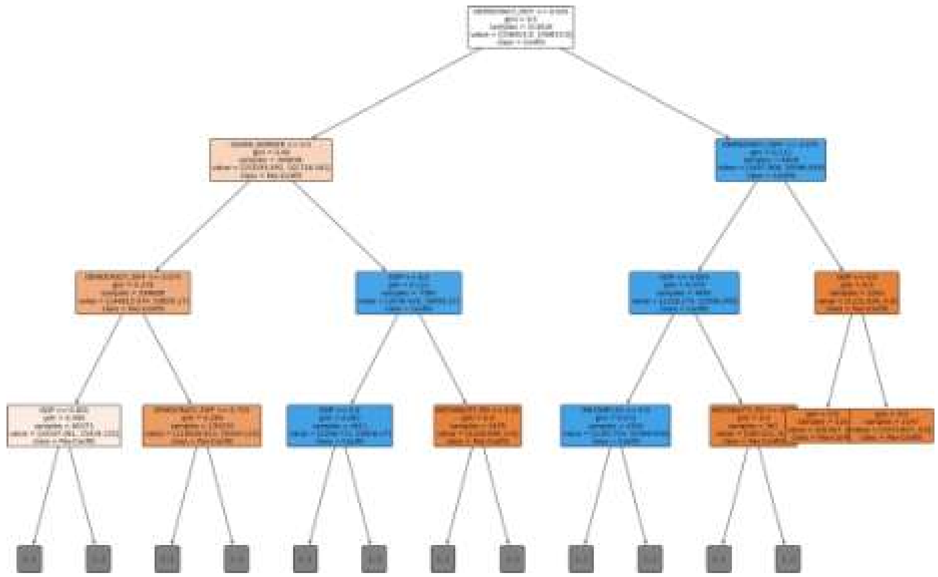

Figure 5. Structure of the Decision Tree

Comparison of Models and Predictors

Table 5 summarizes the different performance metrics for each model, facilitating their comparative evaluation. The study evaluated three machine learning models—Random Forest, Logistic Regression, and Decision Tree to predict interstate armed conflicts in West Africa, focusing on key predictors such as democ- racy level differences, GDP ratios, and shared borders. Among the models, the Random Forest consistently outperformed the others, achieving an accuracy of 98%, a recall of 83%, and an AUC of 80% . The Decision Tree model also demonstrated high recall (95%) but suffered from a lower AUC (53%) due to limitations in handling imbalanced data. Logistic Regression achieved moderate accuracy (79%) but was more prone to false positives, which could compromise its utility in real-world applications.

Table5.PerformanceMetricsofMachineLearningModels

Metric	RandomForest	LogisticRegression	DecisionTree
Accuracy	98%	79%	95%
Precision	100%	100%	100%
Recall	83%	79%	95%
AUC	80%	80%	53%
F1Score	89%	88%	97%

The Random Forest model emerged as the most effective, achieving the highest balance between precision and recall, and demonstrating robust performance even in the presence of class imbalances. Logistic Regression performed moderately well but showed a higher rate of false positives, while the Decision Tree model, though highly accurate, suffered from overfitting, as indicated by its low AUC. When predicting the absence of conflict, reliability is high, particularly for excluding countries not involved in wars. These models are also effective in detecting actual conflicts, as evidenced by their high recall rates (98% and 95%). Moreover, the random forest and logistic regression models show higher AUC scores, indicating remarkable discrimina- tory capability. The F1 scores of the random forest and decision tree models are among the highest, reflecting a perfect balance between precision and recall. Overall, they prove to be the most suitable for our study. As for evaluating the importance of variables in our models, it is essential for judging the success of a machine learning model, which does not rely solely on its precision. In our study on predicting interstate armed conflicts in West Africa, we examined and compared the importance of various features for the two models considered the best (Figure 6).

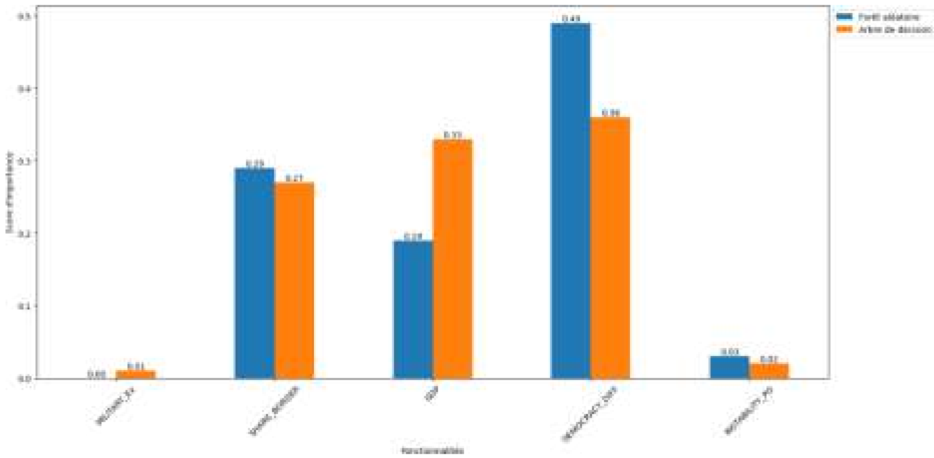

Figure 6. Importance of Explanatory Variables

Thie Figure 6 highlights the significant influence of the divergence in democracy levels between two nations as an indicator of potential armed conflicts. GDP, a macroeconomic aggregate, ranks second, though at a certain distance in the random forest model, indicating that national economic disparities can play a significant role in the emergence of conflicts. Countries with significant economic differences may perceive each other with rivalry, thus generating tensions. Border-sharing has varying importance depending on the presence of other factors. Although internal political instability can lead to conflicts within a country, our models indicate that it is not the dominant factor in interstate armed conflicts. Its impact is relatively minor for all the models examined. However, it can be considered a contributing factor to international tensions. Finally, military expenditures do not weigh as heavily as we might have assumed, even though military power and defense budgets are often associated with a country's tendency to go to war. Nevertheless, this aspect should not be underestimated.

Overview of the West African Case

Our analysis led us to select the random forest model as the best model for predicting the occurrence of interstate armed conflicts. Additionally, the most relevant predictors identified are border-sharing and the difference in democracy levels. With this information, we retrieved the probabilities of interstate conflict events given by the random forest model. We then filtered all dyads sharing common borders. Over the past ten years, for each of these dyads, the maximum probability was selected and presented through Figure 7. From the analysis of this figure, we notice that the highest predicted risks of interstate conflict by our random forest model were for dyads such as Guinea vs. Sierra Leone, Burkina Faso vs. Benin, and Ivory Coast vs. Burkina Faso. These dyads could enter into conflicts if other variables had evolved as predicted by the model. Conversely, the risk of conflict is lower for dyads such as Burkina Faso vs. Togo, Mali vs. Ivory Coast, and Nigeria vs. Niger. There is a significant variation in the probabilities of conflict occurrence among different African dyads. This could be due to several factors not necessarily accounted for in our modeling. We found that differences in democracy levels strongly correlate with the likelihood of interstate armed conflicts. The proposed Random Forest method in this study tended to have an inordinately higher precision (100%) and recall (83%) compared to Logistic Regression and Decision Tree models. Additionally, dyads such as Guinea vs. Sierra Leone and Ivory Coast vs. Burkina Faso exhibited significantly higher predicted conflict probabilities over the past decade, highlighting the role of political and economic disparities in driving regional tensions.

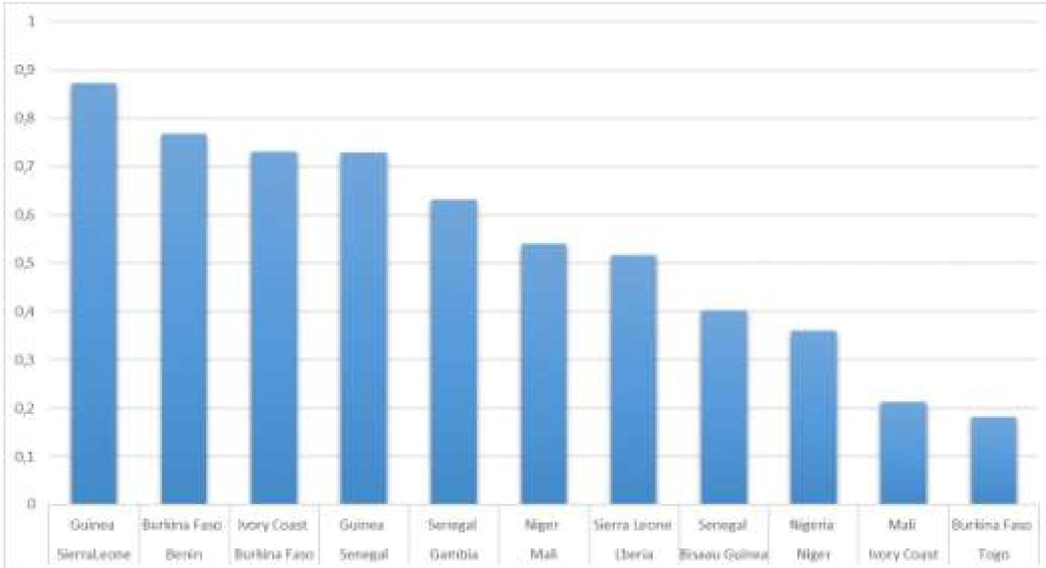

Figure 7. Maximum Probability of Interstate Conflict Occurrence in West Africa Over the Last 10 Years

Discussion

This research highlights the importance of using Machine Learning in conflict prediction. Indeed, among the models used, the random forest model proved to be the best for prediction. This result validates our first hypothesis. Also, the work we have done brings out the key role of the GDP ratio between two nations sharing the same borders in predicting interstate armed conflicts. This confirms our second hypothesis. However, we also noticed that countries with large economic differences may be more likely to experience conflicts. The inclusion of GDP per capita in our analyses can help in understanding this result. Decision- makers and international entities are advised to monitor these economic indicators as possible warning signals of conflicts. The study also confirms the importance of sociopolitical factors, such as changes in the level of democracy and political instability, in predicting conflicts. By exploiting the predictive capacity of models, particularly that of random forest, we can develop early warning systems useful for informing policymakers in West Africa about risk areas, thus allowing preventive interventions. Models based on historical data require regular updating to remain effective in predicting future conflicts. The predictions made allowed us to identify significant variations in the probabilities of conflict occurrence within West African dyads. Some dyads have occurrence probabilities higher than 0.7, as is the case with Ivory Coast vs Burkina Faso. This result confirms our third hypothesis. To address the observed gaps, future research could explore different imputation methods such as model-based imputation or the use of algorithms like KNN to refine datasets and improve model performance. The integration of temporal data could also offer a finer understanding of the temporal dynamics of geopolitical conflicts, thus allowing for better forecasts.

Several studies have explored the use of machine learning for conflict prediction, providing useful comparisons to your findings. For instance, a study in Sub-Saharan Africa found that machine learning mod- els, particularly Random Forest and Gradient Boosting, effectively predicted conflict likelihood when adjusting for class imbalance using techniques like SMOTE. These models highlighted factors such as GDP, proximity to rivers, and past conflicts as key predictors, aligning with your findings that economic disparities and shared borders are significant variables. However, the inclusion of geospatial and environmental factors like nightlight density and rain-fed agriculture in their study adds dimensions that our work has not yet explore [23]. Another study emphasized the role of socioeconomic conditions and governance factors in conflict prediction, under- scoring the need for high-quality data and careful selection of conditioning factors. This resonates with our result that differences in democracy levels are pivotal but suggests expanding our dataset to include governance quality metrics for richer insights [24]. Finally, research focusing on global conflict forecasting suggested that Gradient Boosting models often outperform others in recall, crucial for identifying potential conflict zones. However, Random Forest is noted for its efficiency and accuracy in handling large datasets, consistent with our findings that it outperformed Logistic Regression and Decision Tree models in West Africa. This reinforces the reliability of our methodological choices while suggesting a potential trade-off between speed and model complexity when considering alternatives like Gradient Boosting [25]. These comparisons show that while our study aligns with broader trends in conflict prediction, incorporating additional predictors like environmental factors or testing other algorithms could further enhance our model's predictive power and relevance.

Conclusion.

This study focused on predicting interstate armed conflicts in West Africa, evaluating which machine learning model would be most effective. After comparing several models, including random forest, logistic re- gression, and decision tree, the random forest model stood out for its accuracy and balance between sensitivity and specificity, demonstrating its robustness even in the presence of class imbalances. Our analysis showed that differences in the level of democracy between two countries were a significant predictive factor, corrobo- rating the idea that divergences in political systems increase the risk of conflict. Economic disparities between countries are also recognized for their influence on the probability of conflict occurrence. Although territorial tensions related to shared borders are historically a source of conflict, they were not found to be as determinant as political and economic factors in our study. Similarly, while internal political instability can lead to external conflicts, this factor proved to be the second least influential. Our study found that military capabilities and expenditures were the least influential factor in predicting conflicts. This could be due to how data related to these variables were processed. In terms of perspectives, we envision a more in-depth examination of other data on shared border surfaces, crime index, variations in prices of essential goods, the INFORM risk index, and climate change data to better understand the problem. However, other Machine Learning techniques will be explored for better results that will serve West African decision-makers in developing strategies to reduce the risk of interstate armed conflicts.

This study demonstrates the potential of machine learning models, particularly random forests, to predict interstate armed conflicts in West Africa, emphasizing the importance of factors like border-sharing and differences in democracy levels. However, it falls short of exploring the broader implications of these findings. These results could guide policymakers in designing early-warning systems, focusing resources on mitigating economic disparities and promoting governance reforms. Expanding the analysis to incorporate dynamic or emerging factors such as climate change and refugee flows could improve predictive accuracy. While the models provide valuable insights, the future work should critically examine limitations, such as data imbalances and the ethical challenges of AI in conflict prediction, ensuring robust, actionable outcomes that aid conflict prevention and policy interventions.

References.

[1]H. Nsaibia, "10 conflicts to worry about in 2022: the sahel," 2022.
[2]S. Stocker, "West africa'sgrimtrajectory," 2023.
[3]O. Mateos, J. Renteria, L. TorrellaLlauger et al., "A rethink of the eu'sstrategy in the sahel: fromproblematicpolicy-making to a new 'transformative pragmatism'," 2024.
[4]D. Lewis, "International legalregulation of the employment of artificial-intelligence-related technologies in armedconflict," Moscow Journal of International Law, 2020.
[5]A. Beger, R. K. Morgan, and M. D. Ward, "Reassessing the role of theory and machine learning in forecasting civil conflict," Journal of ConflictResolution, vol. 65, no. 7-8, pp. 1405–1426, 2021.
[6]A. Gnangueˆnon, "Afrique de l'ouest: faire de la pre´vention des conflits la re`gle et non l'exception," Observatoire Boutros-Ghali du maintien de la paix, September, 2018.
[7]M. Luntumbue, "Comprendre la dynamique des conflits: une lecture synthe´tique des facteurs de conflitsenafrique de l'ouest," Groupe de Recherche et d'Information sur la Paix et la Se´curite´ (GRIP), Bruxelles, 2014.
[8]C.-P. David and O. Schmitt, La guerre et la paix: approches et enjeux de la se´curite´ et de la strate´gie. Presses de Sciences po, 2020.

[9]Y. Liu, Y. Wang, and J. Zhang, “New machine learningalgorithm: Randomforest,” in Information Com- puting and Applications: Third International Conference, ICICA 2012, Chengde, China, September 14- 16, 2012. Proceedings 3. Springer, 2012, pp. 246–252.
[10]S. Nusinovici, Y. C. Tham, M. Y. C. Yan, D. S. W. Ting, J. Li, C. Sabanayagam, T. Y. Wong, and C.-Y. Cheng, “Logisticregressionwas as good as machine learning for predicting major chronicdiseases,” Journal of clinicalepidemiology, vol. 122, pp. 56–69, 2020.
[11]H. Hegre, J. Karlsen, H. M. Nyga˚rd, H. Strand, and H. Urdal, “Predictingarmedconflict, 2010–2050,” International StudiesQuarterly, vol. 57, no. 2, pp. 250–270, 2013.
[12]K. Chainoglou, “The protection of intangible cultural heritage in armedconflict: Dissolving the bound-ariesbetween the existinglegalregimes?” Santander Art and Culture Law Review, vol. 3, no. 2, pp. 109–134, 2017.
[13]B. A. Valentino, “Still standing by: whyamerica and the international community fail to preventgenocide and mass killing,” Perspectives on politics, vol. 1, no. 3, pp. 565–578, 2003.
[14]Y. Afina, M. Buchser, A. Krasodomski, J. Rowe, N. Sun, and R. Wilkinson, “Towards a global approach to digital platform regulation,” 2024.
[15]O. Greene and N. Marsh, Small arms, crime and conflict. Taylor & Francis, 2011.
[16]S. Shoker, A. Reddie, S. Barrington, R. Booth, M. Brundage, H. Chahal, M. Depp, B. Drexel, R. Gupta, M. Favaro et al., “Confidence-building measures for artificial intelligence: Workshop proceedings,” arXivpreprint arXiv:2308.00862, 2023.
[17]G. Blainey, Causes of war. Simon and Schuster, 1988.
[18]M. C. . A. Ce´drid and S. Patrick, “Database,” 2024, accessed on 23/07/2024. [Online]. Available: https://drive.google.com/file/d/1e4d 83v599DMlosKyFdg gapjEyzJzAP/view?usp=sharing
[19]M. D. Ward, R. M. Siverson, and X. Cao, “Disputes, democracies, and dependencies: A reexamination of the kantianpeace,” American journal of political science, vol. 51, no. 3, pp. 583–601, 2007.
[20]A. Gelman and J. Hill, Data analysisusingregression and multilevel/hierarchicalmodels. Cambridge universitypress, 2007.
[21]P. Schrodt, P. L. Witt, and A. S. Messersmith, “A meta-analyticalreview of family communication patterns and their associations with information processing, behavioral, and psychosocial outcomes,” Communi- cation monographs, vol. 75, no. 3, pp. 248–269, 2008.
[22]G. L. Cohen and J. Garcia, “Identity, belonging, and achievement: A model, interventions, implications,” Current directions in psychological science, vol. 17, no. 6, pp. 365–369, 2008.
[23]M. Musumba, N. Fatema, and S. Kibriya, “Prevention isbetterthan cure: Machine learningapproach to conflictprediction in sub-saharanafrica,” Sustainability, vol. 13, no. 13, p. 7366, 2021.
[24]T. Obukhov and M. A. Brovelli, “Identifyingconditioningfactors and predictors of conflictlikelihood for machine learningmodels: A literaturereview,” ISPRS International Journal of Geo-Information, vol. 12, no. 8, p. 322, 2023.
[25]H. Hegre, H. M. Nyga˚rd, and P. Landsverk, “Can wepredictarmedconflict? how the first 9 years of publishedforecasts stand up to reality,” International StudiesQuarterly, vol. 65, no. 3, pp. 660–668, 01 2021. [Online]. Available: https://doi.org/10.1093/isq/sqaa094