DATA-DRIVEN ANALYSIS & PREDICTION OF HUMAN RIGHTS VIOLATIONS AGAINST HUMAN RIGHTS DEFENDERS A CASE STUDY: EASTERN AFRICA

ESTHER ASIIMIRE BAGOMBEKA

J23M19/004

A DISSERTATION SUBMITTED TO THE FACULTY OF ENGINEERING, DESIGN AND TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF THE MASTER OF SCIENCE IN DATA SCIENCE AND ANALYTICS OF UGANDA CHRISTIAN UNIVERSITY

October,2025

UGANDA CHRISTIAN UNIVERSITY
A Centre of Excellence in the Heart of Africa

Abstract

Despite the growing availability of big data and machine learning, human rights monitoring in the region remains largely dependent on retrospective reports, eyewitness testimonies, and qualitative assessments, which lack the ability to anticipate future violations. The absence of real-time data processing and predictive analytics limits the ability of policymakers and advocacy groups to implement proactive intervention strategies. As a result, human rights organizations often respond reactively, only after violations occur, rather than deploying preemptive measures to protect HRDs.

In this research, a quantitative research design was adopted, utilizing a cross-sectional approach to analyze patterns in human rights violations. Data was collected from recognized human rights organisations, human rights databases, and global news agencies. The research employed descriptive analytics to identify trends, K-Means clustering to categorize high-risk regions, and predictive modeling to forecast future violations. Seasonal Autoregressive Integrated Moving Average (SARIMA) was used to model long-term seasonal trends, while Recurrent Neural Networks (RNN) captured short-term fluctuations and nonlinear patterns in the data. The Predictive Human Rights Violations Model (PHRVM) emerged as the most effective, balancing structural seasonality and real-time variations, resulting in higher accuracy and improved forecasting reliability compared to individual models.

The findings revealed that human rights violations followed distinct temporal and geographic trends, peaking around election periods, protest seasons, and government crackdowns. While the PHRVM outperformed other forecasting methods during training ($MAE : 0.081$, $RMSE : 0.087$), testing revealed a slight increase in prediction error, with MAE rising to 0.684 and RMSE increasing to 1.109. A paired t-test confirmed that the model significantly outperformed a naïve baseline forecast ($p < 0.05$), validating its predictive capability.

This research concluded that human rights violations follow recognizable patterns, making it possible to anticipate high-risk periods and optimize protection efforts for HRDs. This helps policymakers, and advocacy groups to anticipate risks and implement preventive measures be-

fore violations escalate. The PHRVM's success shows the potential of AI-driven forecasting in social science research, offering a more systematic approach to tracking civic space restrictions. However, for predictive models to be more effective in real-world applications, further refinement is needed, including the integration of real-time data sources such as social media monitoring, remote sensing technologies, and expanded human rights reporting networks. Strengthening these capabilities will enhance model accuracy, responsiveness, and impact, ensuring that human rights organizations can move from reactive responses to preventative protection strategies.

Approval

This is to certify that this research titled "Data - driven analysis and prediction of human rights violations against human rights defenders. A case study: Eastern Africa." has been done under my supervision and is ready for submission.

Supervisor's Name: Mr. John Bosco Wabwire **Signature:** ____________

Acknowledgements

I thank God for the provision, protection, guidance, favour and wisdom during my professional development.

I thank my parents (Eng. Bagombeka Bunnett and Mrs. Bagombeka Charity) for their overwhelming support. I thank my siblings (Arthur, Derrick, Ancel,Vicky and Melon) for their continuous encouragement. I thank my friend (Brendan) for moving closely with me through this journey and always reminding me why I started this journey.

I would like to thank my supervisor Mr. John Wabwire for his dedicated support and guidance. He has continuously provided encouragement and was always willing and enthusiastic to assist in any way he could throughout the research project.

May the Lord God Bless you all abundantly!

Declaration

I, Bagombeka Esther Asimiire, hereby declare that the research titled "Data driven analysis of trends and Predicting Of Human Rights Violations Against Human Rights Defenders.A case study of Eastern Africa" is my original work, and as far as currently known, it has never been submitted for consideration for any kind of academic award to any university or other institution. I affirm that all sources used in this proposal have been properly acknowledged and cited. I understand the consequences of plagiarism and academic dishonesty, and I attest to the integrity and authenticity of this work.

Candidate's Name: Bagombeka Esther Asiimire **Signature:** ____________________

Table of Contents

Abstract . . . . **i**

Approval . . . . **iii**

Acknowledgements . . . . **iv**

Declaration . . . . **v**

Table of Contents . . . . **vi**

List of Tables . . . . **x**

List of Figures . . . . **xi**

List of Acronyms . . . . **xiii**

1 Introduction **1**

1.1 Overview 1

1.2 Background 2

1.2.1 Historical Background 2

1.2.2 Conceptual Background 3

1.2.3 Contextual Background 4

1.3 Problem Statement 5

1.4 Research Objectives 6

1.5 Research Questions 6

1.6 Hypotheses 7

1.7 Contribution to research 7

1.8 Scope of study 8

2 Literature Review **9**

2.1 Conceptual Framework 9

2.1.1 Theoretical Framework 10

2.2 Prior Empirical studies 12
2.2.1 Predicting Legal Decisions in Human Rights Cases Methodology . . . 13
2.2.2 Automated Data Pipelines for Human Rights Analytics 14
2.2.3 Predicting Attacks Against Human Rights Defenders 14
2.2.4 Machine Learning for Human Rights Monitoring in Africa 15
2.2.5 Remote Sensing and Machine Learning for Human Rights Monitoring Methodology 16
2.2.6 NLP and Crowdsourcing for Human Rights Monitoring 17
2.2.7 Limitations of Traditional Human Rights Monitoring: Insights from Amnesty International and Human Rights Watch 18
2.2.8 Summary of results from some of the studies reviewed 18

3 Methodology **21**
3.1 Philosophical Paradigm 21
3.2 Study Site 22
3.3 Research Design 23
3.4 Area of Study 24
3.5 Study Population 25
3.5.1 The Inclusion Criteria 25
3.5.2 The Exclusion Criteria 25
3.6 Source of Information 25
3.7 Variables definitions 26
3.8 Quality/Error Control 28
3.9 Data processing and analysis 29
3.9.1 Descriptive Analysis 29
3.9.2 Clustering Analysis with K-Means 30
3.9.3 Predictive Modeling 30
3.9.4 Model Evaluation and Hypothesis Testing 30

3.10 Ethical considerations 31

4 Data Analysis, Presentation and Interpretation of Findings 33

4.1 Data Cleaning and Preprocessing 33

4.1.1 Handling Missing Values 33

4.1.2 Standardizing Categories 34

4.2 Analyze patterns of human rights violations against HRDs in the Eastern Africa, including the key threats, targeted groups, and trends over time. 34

4.2.1 Distribution of Violation Types by Country 35

4.2.2 Geographic Distribution of Human Rights Violations 36

4.2.3 Key threats against human rights defenders 36

4.2.4 Human rights violations 38

4.2.5 Seasonality Analysis of Human Rights Violations 51

4.2.6 Clustering Analysis of Human Rights Violations 52

4.3 To design a predictive modeling framework that integrates statistical and machine learning techniques for forecasting human rights violations. 56

4.3.1 SARIMA Model 56

4.3.2 Predictive Human Rights Violations Model (PHRVM) 63

4.4 To evaluate the accuracy and performance of the predictive model, testing its effectiveness in forecasting human rights violations using appropriate validation metrics. 66

4.4.1 Hypothesis Testing 67

4.4.2 Areas Where the Model Underperformed 69

4.5 Methodological constraints/bias 70

5 Discussion 72

5.1 Patterns and Clustering of Human Rights Violations in Eastern Africa 72

5.2 Performance and Forecasting Insights of the PHRVM Model 75

6 Conclusion and Recommendations 77

6.1 Conclusion 77

6.2 Recommendations 78

6.2.1 Create a vast database for human rights violations against HRDs 78

6.2.2 Strengthening Data Collection and Real-Time Monitoring 79

6.2.3 Integration of Predictive Insights into HRD Protection Strategies 79

6.2.4 Refining the Predictive Model for Higher Accuracy 80

6.3 Future Research 80

List of Tables

Table	Description	Page
Table 1	Summary of Studies Using Data Science in Human Rights Research	20
Table 2	Organizations and Platforms Used for Data Collection	26
Table 3	Human Rights Databases Used for Data Collection	26
Table 4	Variables Included in the Human Rights Dataset	28
Table 5	Distribution of Rights Violations Across Clusters	55
Table 6	Median Values for Each Type of Violation by Cluster	55

List of Figures

Figure	Description	Page
Figure 1	Conceptual and analytical framework for PHRVM	31
Figure 2	Pivot table showing the amount of civic violations across the region	35
Figure 3	Horizontal bar chart showing the frequency of violations across the region	36
Figure 4	Pie chart showing the top 10 distribution of threats from the period 2020-2023	37
Figure 5	Bar chart showing the distribution civic violations in 2020	38
Figure 6	Horizontal bar chart showing the top 15 groups violated in 2020	39
Figure 7	Horizontal bar chart showing top 15 threats faced by HRDs in 2020	40
Figure 8	Bar chart showing the distribution civic violations in 2021	42
Figure 9	Horizontal bar chart showing the top 15 groups violated in 2021	43
Figure 10	Horizontal bar chart showing the top 15 threats violated in 2021	44
Figure 11	Distribution of violations by country in 2021	45
Figure 12	Bar chart showing the distribution civic violations in 2022	46
Figure 13	Horizontal bar chart showing the top 15 groups violated in 2022	46
Figure 14	Horizontal bar chart showing top 15 threats faced by HRDs in 2022	47
Figure 15	Distribution of violations by country in 2022	48
Figure 16	Bar chart showing the distribution civic violations in 2022	49
Figure 17	Horizontal bar chart showing the top 15 groups violated in 2023	49
Figure 18	Horizontal bar chart showing top 15 threats faced by HRDs in 2023	50
Figure 19	Distribution of violations by country in 2023	51
Figure 20	Seasonality distribution for Freedom of Expression seasonality violations	52
Figure 21	Seasonality distribution for Freedom of Association violations	52
Figure 22	Seasonality distribution for Freedom of Peaceful Assembly violations	52
Figure 23	Elbow method for optimal number of clusters according to the violations	54

Figure 24 Area graph showing ACF Plot 58
Figure 25 Area graph showing ACF Plot 58
Figure 26 SARIMA model accuracy test 61
Figure 27 RNN model predictions 63
Figure 28 PHRVM predictions for violations in 2024 66
Figure 29 Actual vs Predicted violations during training in 2024 66
Figure 30 PHRVM Model predictions for violations in 2024 alongside the actual violations 67

List of Acronyms

ACF	Autocorrelation Function
ACLED	Armed Conflict Location & Event Data Project
AI	Artificial Intelligence
BBC	British Broadcasting Corporation
CNN	Cable News Network
CIVICUS	World Alliance for Citizen Participation
CPJ	Committee to Protect Journalists
CSO	Civil Society Organization
ECHR	European Court of Human Rights
HRD	Human Rights Defender
ICCPR	International Covenant on Civil and Political Rights
KNN	K-Nearest Neighbors
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MSE	Mean Squared Error
NGO	Non-Governmental Organization
NLP	Natural Language Processing
OHCHR	Office of the High Commissioner for Human Rights
PACF	Partial Autocorrelation Function

PHRVM	Predictive Human Rights Violations Model
p-value	Probability Value
RNN	Recurrent Neural Network
RMSE	Root Mean Squared Error
RSF	Reporters Without Borders (Reporters Sans Frontières)
SARIMA	Seasonal Autoregressive Integrated Moving Average
SDG	Sustainable Development Goal
T-test	Statistical test to compare means
UDHR	Universal Declaration of Human Rights
UNHRC	United Nations Human Rights Council

Chapter 1

Introduction

1.1 Overview

Data science has become an important asset in human rights research, offering advanced methodologies to systematically analyze and predict violations. By employing machine learning algorithms and predictive analytics, researchers can process extensive datasets to identify patterns and forecast potential abuses.Despite the increasing availability of big data, machine learning, and predictive analytics, the systematic application of data science in human rights research remains underdeveloped, particularly in the context of protection of human rights defenders (HRDs) in Africa. Previous studies have demonstrated the potential of data science for analyzing human rights violations, such as Marazzi (2016), who explored automated data pipelines for processing large-scale human rights data, and Pandey et al. (2018), who employed convolutional neural networks (CNNs) for detecting patterns of violence from imagery and text sources. However, these methodologies have not been widely applied to forecast human rights violations against HRDs using structured historical data and advanced time-series modeling techniques. The lack of predictive analytics in this domain limits the ability of human rights organizations and policymakers to anticipate risks and implement proactive interventions.

This research addresses this gap by using data science methodologies to systematically analyze and forecast human rights violations against HRDs in Eastern Africa. Using historical data from 2020 to 2023, the study applies machine learning algorithms, time-series forecasting techniques (such as SARIMA and Recurrent Neural Networks), and geospatial analysis to uncover trends and patterns in violations. By integrating these advanced analytical techniques, the study aims to develop a predictive model that can forecast the probability of violations in the region.

The application of data science to human rights research represents a transformative shift, allowing for predictive modeling of violations that would otherwise go undocumented or unanticipated. By integrating statistical learning, machine learning, and time-series forecasting models, this study helps in the understanding of the temporal and spatial dimensions of violations but also contributes to improve the protection and safety of HRDs. The findings from this research will support human rights advocacy efforts, policy formulation, and strategic intervention planning by equipping stakeholders with data-driven insights to anticipate and respond to human rights threats with greater precision and efficiency.

1.2 Background

1.2.1 Historical Background

The Evolution of Human Rights Monitoring

Historically, human rights monitoring has depended on qualitative documentation through reports, witness testimonies, and legal advocacy. Organizations such as Amnesty International, Human Rights Watch, and the United Nations have played a central role in documenting violations. However, traditional approaches have faced significant challenges, including delayed reporting, under documentation in repressive regimes, and the challenge of large-scale data management (Amnesty International, 2023; Human Rights Watch, 2024). The 21st century introduced big data analytics and computational methodologies, which have transformed the way human rights violations are tracked and analyzed. Techniques such as geospatial mapping, natural language processing (NLP), and machine learning algorithms allow for real-time monitoring, trend identification, and risk prediction (Medvedeva et al., 2020).

Despite the growing application of data-driven methodologies in human rights research, predictive analytics remains underutilized in Africa, particularly in forecasting threats against human rights defenders (HRDs). While studies such as Marazzi (2016) and **Pandey2018a<empty citation>** have successfully applied machine learning models to analyze social unrest and violence globally, similar techniques have not been fully adapted to regions with limited data availability, censorship, and systemic state repression. This study seeks to fill this research gap by applying predictive modeling techniques to analyze and forecast human rights violations against HRDs in Eastern Africa.

1.2.2 Conceptual Background

Key Terms and Analytical Constructs

The concept of human rights defenders (HRDs) refers to individuals or groups actively advocating for fundamental freedoms, such as freedom of speech, association, and assembly. HRDs include journalists, activists, legal professionals, and whistleblowers, who frequently face persecution for their work (Forst, 2018). Their role in promoting transparency, holding governments accountable, and defending marginalized communities makes them particularly vulnerable to state repression. Governments and non-state actors often target HRDs through arbitrary arrests, judicial harassment, surveillance, and physical violence (Civicus, 2023).

Civic space is another crucial concept in human rights research, referring to the degree of freedom individuals and organizations have to engage in public discourse, advocacy, and activism. In many authoritarian regimes, civic space is systematically restricted through censorship, legal barriers, and forceful crackdowns on civil society organizations (CSOs) (Reporters Without Borders, 2023). Shrinking civic space directly correlates with increasing human rights violations, as governments deploy various mechanisms to control public dissent (Smith, 2020).

In the era of big data, data science has become a powerful tool for human rights research. Predictive analytics enables the identification of risk patterns and emerging threats, improving early warning systems and targeted intervention strategies. Key data science techniques used in human rights analysis include natural language processing (NLP) for extracting critical information from news reports, geospatial analysis for mapping the distribution of human rights violations, and time-series forecasting models (SARIMA and RNNs) for predicting future threats (Alhelbawy et al., 2020; Medvedeva et al., 2020). By integrating these computational methodologies, this study seeks to enhance the ability of human rights organizations to anticipate risks and implement data-driven interventions to protect HRDs.

1.2.3 Contextual Background

Human Rights Violations Against HRDs in Eastern Africa

The Eastern Africa region is characterized by a high prevalence of human rights violations against HRDs, with state and non-state actors actively suppressing civic freedoms. According to DefendDefenders (2023), Amnesty International (2023),Civicus (2023), HRDs in this region face systematic repression through arbitrary arrests, surveillance, and physical violence. In some countries, legal frameworks are deliberately weaponized to criminalize activism, restricting freedom of association, expression, and peaceful assembly.

Countries such as Uganda, Ethiopia, and Kenya exhibit cyclical patterns of repression, where HRDs experience heightened threats during election periods, protests, and political crises (Human Rights Watch, 2024). Meanwhile, nations such as Eritrea, Djibouti, and Rwanda maintain long-term restrictions on civic space, where HRDs operate under strict surveillance, censorship, and state-controlled legal systems. The nature of violations varies across the region, with extrajudicial killings and enforced disappearances more prevalent in conflict-affected states such as Somalia and South Sudan, while digital repression and media censorship are widespread in countries with more technologically advanced authoritarian regimes (Civicus, 2023).

Despite extensive documentation of human rights abuses, the region lacks systematic forecasting methods to predict when and where HRDs are most at risk. Current human rights reports tend to be retrospective, focusing on documenting past violations rather than anticipating future risks. This study seeks to address this gap by developing a predictive model that integrates historical trends, geographic data, and real-time monitoring to improve HRD protection strategies. By applying data science techniques such as machine learning, time-series forecasting, and clustering analysis, this research will contribute to evidence-based advocacy and strategic intervention planning for HRDs in Eastern Africa.

1.3 Problem Statement

In Eastern Africa, the increasing availability of big data, machine learning, and predictive analytics has transformed fields such as healthcare, finance, and security, yet human rights monitoring remains heavily reliant on traditional qualitative methods. Currently, approaches such as eyewitness testimonies, retrospective reports, and case documentation, provide valuable historical insights but lack predictive capabilities to anticipate emerging threats. The absence of real-time data processing, pattern recognition, and forecasting models limits the ability of policymakers and advocacy groups to proactively mitigate risks and intervene before violations occur.

Prior research has emphasized the need for data-driven approaches in human rights monitoring, yet predictive analytics remains underutilized in African contexts (Dancy & Fariss, 2017).Despite the proven success of AI and machine learning in detecting patterns of violence, previous studies, such as Marazzi (2016) and Medvedeva et al. (2020), have demonstrated the potential of artificial intelligence in human rights research, using machine learning to analyze judicial decisions and identify patterns of violence. However, these methodologies have not been widely applied to HRDs in the Eastern Africa, particularly in developing predictive models that can forecast when and where violations are most likely to occur.

To address this gap, this research aims to develop a machine learning-based predictive model to forecast human rights violations against HRDs in Eastern Africa, using historical data from 2020 to 2023. The study will utilize time-series forecasting techniques (SARIMA), deep learning models (RNN), and geospatial analysis to uncover patterns, trends, and risk factors associated with these violations and finally develop a Predictive Human Rights Violations Model (PHRVM.The model's performance will be evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to ensure its reliability in predicting future violations.

The findings from this research will offer practical implications for human rights organizations, policymakers, and advocacy groups. By using data-driven forecasting, stakeholders can anticipate high-risk periods, implement targeted interventions, and strengthen protective measures for HRDs. It also enables shifting from a reactive approach to a proactive strategy that improves the safety and resilience of HRDs across Eastern Africa.

1.4 Research Objectives

Main Objective

To develop an effective predictive model for forecasting human rights violations against HRDs in Eastern Africa based on historical data from 2020 to 2023.

Specific Objectives

1. To analyze patterns of human rights violations against HRDs in the Eastern Africa, including the key threats, targeted groups, and trends over time.

2. To identify and select suitable machine learning and statistical techniques for designing a predictive model for forecasting human rights violations.

3. To evaluate the accuracy and performance of the predictive model, testing its effectiveness in forecasting human rights violations using appropriate validation metrics.

1.5 Research Questions

1. What are the key threats and groups targeted in human rights violations against HRDs in the Eastern Africa from 2020 to 2023?

2. Can a hybrid predictive model (PHRVM) that integrates statistical and machine learning techniques be effectively used to forecast human rights violations against HRDs based on historical data?

3. How accurate and effective is the Predictive Human Rights Violations Model in forecasting human rights violations against HRDs based on historical data?

1.6 Hypotheses

1. Null Hypothesis (Ho): There is no significant relationship between historical patterns of human rights violations and the ability to accurately predict future violations against HRDs in the Eastern Africa.

2. Alternative Hypothesis (H1): Historical patterns of human rights violations can be used to develop a predictive model that accurately forecasts future violations against HRDs in the Eastern Africa.

1.7 Contribution to research

This research contributes to the existing knowledge base surrounding human rights violations against HRDs in Eastern Africa and the application of data science methodologies in human rights research.

The development of a predictive model for forecasting violations against HRDs provides valuable insights for future research endeavors and policy interventions aimed at protecting HRDs and promoting human rights in Eastern Africa. Furthermore, the practical implications of this study extend to advocacy efforts and policy formulation, as stakeholders can utilize the findings to better tailor their interventions to address the specific needs of HRDs and enhance their safety and security.

Moreover, this research directly contributes to Sustainable Development Goal 16 (SDG16) by promoting peace, justice, and strong institutions. Methodologically, this study also contributes to the refinement and validation of machine learning techniques for human rights research, enhancing the reliability and applicability of predictive models in the protection of HRDs. Overall, the findings of this study deepen our understanding of human rights violations against HRDs in Eastern Africa and provide actionable insights for stakeholders working towards the protection and promotion of human rights in the region.

1.8 Scope of study

The scope of this study is designed to address a critical gap identified in the literature: the need for a data-driven, predictive approach to human rights violations in the Eastern Africa. While existing studies and reports provide in-depth accounts of rights violations and threats faced by specific groups, they lack systematic forecasting methods to anticipate future risks. By focusing on countries such as Uganda, Ethiopia, Somalia, Kenya, and others within this region, the study targets areas with complex socio-political landscapes where violations are recurrent yet underrepresented in predictive research.

The chosen timeframe of 2020 to 2023 supports this goal by incorporating recent trends in violations to provide a solid historical foundation for prediction. This period is particularly relevant as it reflects the impact of recent policy shifts, socio-political tensions, and regional conflicts, all of which contribute to the dynamics affecting civic space. By focusing on recent data, the study aims to improve the accuracy and relevance of its forecasting model, filling the current research gap with a timely, data-driven approach.

This scope is also informed by the need for a specific approach that prioritizes forecasting techniques over traditional social science analysis. Using methods such as time series analysis and recurrent neural networks (RNNs), the study aligns with data science methodologies, positioning itself as a valuable tool for organizations and policymakers seeking to proactively address civic space threats.

Chapter 2

Literature Review

2.1 Conceptual Framework

This research integrates data-driven analysis, clustering techniques, and predictive modeling to examine human rights violations against HRDs in the Eastern Africa. The framework aligns with the objectives by structuring data collection and analysis methods around key research variables, while also using predictive techniques to anticipate future violations.

The primary variables used in data collection include case details (violation type, date, location), affected groups (journalists, activists, opposition members, protesters), threats on HRDs (arbitrary arrests, judicial harassment, censorship, killings), and contextual factors or seasonality (election cycles, government policies, protests, and crackdowns). These variables were manually extracted from reputable sources that include human rights reports, news articles, and human rights databases, ensuring a structured dataset for analysis. Data preprocessing involved removing redundancies, handling missing values, and standardizing categorical variables to maintain accuracy and reliability.

In the analysis phase, different methodologies were applied based on the study's objectives. To identify patterns of human rights violations (Objective 1), analysis was conducted to determine the most common violations and at-risk populations. This approach allowed for a quantitative assessment of trends across different years and locations to understand how threats evolved. To analyze temporal and geographic trends K-Nearest Neighbors (KNN) clustering was used to classify countries into distinct groups based on similarity in violation patterns. For Objective 2, which focuses on developing a predictive modeling framework, a PHRVM model integrating SARIMA (Seasonal Autoregressive Integrated Moving Average) and Recurrent Neural Networks (RNN) networks was used. SARIMA captured seasonal and long-term trends in human rights violations, particularly in relation to political events such as elections and government-imposed crackdowns. The RNN model, on the other hand, was used to analyze complex, non-linear relationships in the data, capturing patterns that traditional time series models may overlook. Using the hybrid modeling approach enhanced forecasting accuracy.For objective 3, the model will be tested using the Mean Average Error (MAE) and Root Mean Squared Error (RMSE) as well as a hypothesis test will be done using the naïve baseline model

While this study primarily follows a conceptual framework based on data science methodologies, it also draws from theoretical perspectives relevant to human rights violations and civic space analysis. The research aligns with political repression theory, which posits that governments restrict civic space through coercive means, particularly in response to perceived threats from opposition groups and activists (Earl, 2003). Additionally, elements of social movement theory are considered, as patterns of human rights violations often correlate with protest cycles, government reactions, and civil resistance movements (Alimi, 2009). These insights help explain the underlying motivations for human rights violations and provide a contextual foundation for the predictive models.

The combination of quantitative analysis and predictive modeling ensures that findings can inform evidence-based advocacy, policy recommendations, and strategic interventions aimed at safeguarding HRDs across the Eastern Africa. This helps bridge the gap between data science and human rights protection by offering practical tools for early warning systems and proactive human rights defense strategies.

2.1.1 Theoretical Framework

In analyzing and forecasting human rights violations in the Eastern Africa, several advanced analytical techniques could theoretically achieve the study's objectives. However, each alternative method comes with distinct advantages and limitations. A comparative analysis of Decision Trees and Random Forests, Support Vector Machines (SVM), Autoregressive Integrated Moving Average (ARIMA), SARIMA, and Long Short-Term Memory (LSTM) networks reveals why KNN clustering, RNN, and Seasonal Autoregressive Integrated Moving Average (SARIMA) were ultimately chosen for this study.

Decision Trees and Random Forests are popular methods in classification, valued for their interpretability and robustness in handling diverse data types. Decision Trees, and their ensemble form in Random Forests, can classify regions or threats by breaking down complex data into hierarchical rules (Breiman, 2001). They could theoretically group countries or types of violations, thus identifying risk patterns for human rights defenders (HRDs). However, these models lack the capacity to process time-dependent data. They are more effective for static classification, rather than for forecasting, as they cannot capture trends or sequential dependencies.

Support Vector Machines (SVM) are another commonly used classification tool, capable of separating data classes in high-dimensional spaces by maximizing the margin between different clusters (Cortes & Vapnik, 1995). For this study, SVM could categorize data, such as identifying which violations correlate with specific target groups. Yet, SVM's applicability to time-dependent data is limited, as it focuses on point-in-time classifications rather than handling multivariate, sequential datasets. Additionally, SVMs are computationally intensive, which poses practical challenges given the large-scale nature of human rights data.

On the other hand, K-Nearest Neighbors (KNN) offers a non-parametric, instance-based learning approach that makes predictions for new data points based on their proximity to known previous cases. It is most beneficial in the sense that it is able to detect similar patterns along both time and feature dimensions, and hence making it more adaptable to subtle shifts. KNN is also simpler to calculate and interpret, which makes it a good option to be used in early warning systems.

ARIMA is a widely recognized model in time-series analysis, relying on autoregressive and moving-average components to capture linear patterns over time (Box & Jenkins, 1976). ARIMA could provide a simple yet effective method for forecasting trends in violation frequencies by analyzing historical data. However, its primary limitation is its focus on linear relationships, which restricts its utility in this study. Human rights violations are driven by complex, non-linear factors, such as political and social conditions, that ARIMA cannot fully capture. While SARIMA, an extension of ARIMA, improves on this by incorporating seasonality, it still struggles to model unpredictable, non-linear variations.

LSTM neural networks, a variant of RNNs, have gained traction in sequence prediction, particularly for long-sequence time-series forecasting (Hochreiter & Schmidhuber, 1997). LSTM networks excel at capturing long-term dependencies due to their memory cells, which are highly effective for prolonged time-series data. While LSTM could theoretically improve the model's predictive power, it is computationally complex and more suitable for long-term data series. Given this study's shorter temporal span (2020-2023), standard RNNs provide sufficient accuracy and are computationally more efficient for the timeframe and data complexity, making them a more appropriate choice.

Given the limitations of standalone models, a hybrid approach combining SARIMA and RNN was adopted in this study to complement strengths of both statistical and deep learning models.SARIMA was incorporated to model long-term trends and seasonal patterns in human rights violations. Its ability to capture seasonal fluctuations makes it particularly useful in detecting cyclical repression patterns linked to elections, government crackdowns, and civic unrest—events that historically lead to increased human rights violations in the Eastern Africa.However, SARIMA alone cannot capture the non-linear, dynamic nature of violations driven by political and social upheavals. To address this, RNN was integrated to model non-linear dependencies and detect unexpected shifts in violations, such as sudden policy changes, mass protests, or security crackdowns. Unlike SARIMA, RNN can learn from past sequences and adjust predictions based on recent trends, making it better suited for forecasting unexpected surges in violations.

2.2 Prior Empirical studies

This section reviews existing research on the application of machine learning and predictive analytics in human rights violations, with a specific focus on methodologies, datasets, findings, and recommendations. While various studies have explored pattern recognition and risk assessment, few have attempted to forecast human rights violations using predictive models.

2.2.1 Predicting Legal Decisions in Human Rights Cases Methodology

Medvedeva et al. (2020) explored the use of machine learning in human rights research by developing models to predict judicial decisions of the European Court of Human Rights (ECtHR). The study applied Support Vector Machines (SVM), Random Forest, and Logistic Regression, using Natural Language Processing (NLP) techniques to extract information from legal texts. The dataset consisted of over 1,000 past ECtHR cases, sourced from HUDOC, the ECtHR's online database of court rulings. The results showed that SVM was the best-performing model, achieving an accuracy of 79%, demonstrating that machine learning can be an effective tool for predicting judicial outcomes.

Despite these promising results, the study had significant limitations. It focused solely on classification tasks, predicting whether a court case would result in a conviction or acquittal, rather than developing a model to forecast human rights violations before they occur. The authors acknowledged this gap and recommended that future research expand machine learning applications beyond legal case classification to broader human rights violations. Their findings highlight the potential of machine learning in human rights research but also underscore the need for predictive analytics, an area that this study seeks to address by developing a forecasting model for human rights violations against HRDs.

This research extends the work of Medvedeva et al. by applying machine learning and statistical forecasting methods to predict violations in real-world civic space, specifically targeting risks to HRDs in Eastern Africa. It introduces temporal modeling approaches to address sequential dependencies that prior classification-focused studies do not capture.

2.2.2 Automated Data Pipelines for Human Rights Analytics

Marazzi (2016) investigated the application of machine learning in human rights analytics by developing an automated data pipeline to process large-scale human rights data. The study employed unsupervised learning techniques to recognize patterns in violations but did not incorporate forecasting models such as time-series analysis. The dataset included social unrest and human rights reports from major NGOs like Amnesty International and Human Rights Watch, along with publicly available data on human rights violations between 2010 and 2015.

The findings demonstrated that machine learning could successfully classify human rights violations into categories such as arbitrary detention and forced disappearances. Additionally, the study identified government policies and political transitions as significant predictors of human rights abuses. However, a major limitation was its focus on real-time data processing rather than future prediction. Recognizing this gap, Marazzi (2016) recommended that future research explore predictive analytics to anticipate potential violations rather than solely categorizing past incidents.

Although this study highlights the importance of machine learning in human rights analysis, it does not explore how to anticipate when and where future violations might occur. Our study builds directly on this gap by shifting from classification to prediction — forecasting future violations using structured historical data and advanced time-series models. By integrating both seasonal statistical techniques and non-linear neural models , this study introduces a novel predictive framework (PHRVM) that enhances the applicability of machine learning in HRD protection.

2.2.3 Predicting Attacks Against Human Rights Defenders

Ran et al. (2023) explored the use of Natural Language Processing (NLP) and machine learning classifiers to analyze attacks against Human Rights Defenders (HRDs). The study applied decision trees, random forests, and deep learning models to classify the severity of violations based on textual data. However, while the study successfully categorized different types of attacks, it did not incorporate time-series forecasting to predict when future violations might occur. The dataset consisted of news articles and human rights reports spanning over five years, compiled from various NGO sources and labeled to reflect documented attacks on HRDs.

The results demonstrated that the random forest model performed best, achieving 85% accuracy in classifying attacks. The study also identified key risk factors that increased the likelihood of HRD violations, including government crackdowns and periods of public protest. However, its primary limitation was its focus on classification rather than prediction. Recognizing this gap, the authors recommended that future research develop forecasting models capable of predicting when and where attacks on HRDs would occur. While this study highlights the feasibility of applying machine learning to human rights violations, it lacks predictive capabilities, a gap that this research aims to fill by integrating time-series forecasting and geospatial analysis to anticipate future human rights violations against HRDs.

2.2.4 Machine Learning for Human Rights Monitoring in Africa

Smith (2020) examined the use of machine learning techniques for human rights monitoring in Africa, applying unsupervised learning methods such as K-Means clustering and Latent Class Analysis to group countries based on patterns of human rights violations. Additionally, the study incorporated geospatial analysis to map human rights trends and visualize areas with high levels of repression. However, while the study successfully identified regional patterns of human rights abuses, it did not develop predictive models to forecast future violations.

The dataset used in the study included the CIRIGHTS dataset, which compiles global human rights data from 1981 to 2019 across more than 150 countries, along with reports from Amnesty International and Human Rights Watch. The findings revealed that elections and political instability strongly correlated with increased human rights violations, particularly in African nations. By clustering countries with similar patterns of repression, the study provided insights into which regions were most at risk. However, its major limitation was its lack of forecasting capabilities, as it focused on grouping countries based on past violations rather than predicting future risks.

Recognizing this gap, Smith (2020) recommended that future research integrate time-series forecasting models to predict when and where HRD violations might escalate. While this study supports the use of clustering and geospatial analysis in human rights research, it does not provide a predictive framework. This research builds on Smith's findings by incorporating time-series forecasting techniques (SARIMA, RNN models) and machine learning-based risk assessment, allowing for the prediction of human rights violations rather than just the identification of historical patterns.

2.2.5 Remote Sensing and Machine Learning for Human Rights Monitoring Methodology

Quinn et al. (2018) applied remote sensing techniques combined with machine learning algorithms to detect human rights violations in conflict zones. They analyzed satellite imagery using computer vision models to identify destroyed structures and damaged settlements, which were indicative of attacks on civilian populations. Supervised classification models were trained on labeled imagery datasets to differentiate between normal landscapes and those affected by human rights abuses.

The study utilized high-resolution satellite images from sources such as Google Earth, NASA's Landsat program, and commercial satellite providers. The dataset covered multiple conflict-affected regions, focusing on areas with known violations documented by human rights organizations.The machine learning models successfully identified destruction patterns, achieving high accuracy in detecting areas impacted by human rights violations.The study demonstrated that remote sensing could serve as an early warning tool, especially in regions with limited on-the-ground reporting.

This approach was limited to detecting large-scale physical destruction and could not capture less visible human rights violations, such as arrests, judicial harassment, or digital repression.

2.2.6 NLP and Crowdsourcing for Human Rights Monitoring

Alhelbawy et al. (2020) developed Ceasefire Iraq, an NLP-based platform that monitored human rights violations through social media data and crowdsourced reports. The study applied Natural Language Processing (NLP) models, including sentiment analysis and entity recognition, to classify posts related to human rights abuses. Supervised learning algorithms, such as Support Vector Machines (SVM) and Recurrent Neural Networks (RNNs), were used to detect and categorize violations.

In this study,Twitter API was used to collect real-time tweets related to human rights violations.The dataset included crowdsourced reports from local human rights organizations and eyewitness testimonies. Text data was pre-processed using tokenization, stopword removal, and word embedding techniques (Word2Vec and BERT) to enhance classification accuracy.

The NLP model successfully categorized violations, allowing real-time tracking of state repression, armed conflict, and civilian casualties.Crowdsourcing improved data availability, particularly in regions with restricted press freedom.However, the study focused on the Middle East, and the NLP models required language adaptation before they could be applied in African regions. Additionally, social media monitoring faces bias, as governments may manipulate information or censor content.

Our study addresses these limitations by applying structured predictive modeling techniques to historical event data, rather than relying on unstructured real-time social media input. In doing so, we introduce a forecasting dimension to human rights monitoring, offering tools for proactive intervention. This work contributes to the expansion of data science methods in human rights research by demonstrating how hybrid models can predict civic space threats beyond real-time classification.

2.2.7 Limitations of Traditional Human Rights Monitoring: Insights from Amnesty International and Human Rights Watch

Amnesty International (2023) and Human Rights Watch (2024) produce annual reports documenting human rights violations globally, with a particular focus on Human Rights Defenders (HRDs) in the Eastern Africa. Their methodology relies primarily on qualitative analysis, collecting eyewitness accounts, legal documentation, and government reports to identify patterns of repression. Unlike machine learning-based approaches, these organizations use manual classification of violations, making their analysis descriptive and retrospective rather than predictive.

The datasets compiled in these reports come from NGOs, journalists, and international bodies, offering extensive documentation of human rights violations across different regions. While these reports provide valuable insights into patterns of HRD persecution, they do not incorporate data science techniques or predictive modeling to anticipate future violations. The main limitation of these reports is their inability to forecast human rights risks, as they primarily focus on documenting past incidents rather than predicting future trends.

Recognizing this shortcoming, both organizations have emphasized the need for integrating data science and predictive analytics into human rights research. They encourage the use of technology to enhance human rights monitoring and risk assessment. This study directly addresses these recommendations by applying machine learning-based forecasting models to predict when and where human rights violations against HRDs are likely to occur, transitioning from retrospective documentation to proactive forecasting.

2.2.8 Summary of results from some of the studies reviewed

The reviewed literature demonstrates growing interest in the application of machine learning, natural language processing, clustering, and geospatial analysis in the field of human rights monitoring. Studies such as Medvedeva et al. (2020) and Ran et al. (2023)successfully applied ML classifiers to predict legal outcomes or categorize HRD violations. Others like Smith (2020) and Marazzi (2016)showed the value of clustering and automated pipelines in understanding regional repression patterns. Remote sensing work by Quinn et al. (2018) and NLP platforms like Ceasefire Iraq Alhelbawy et al. (2020) further highlighted how advanced computational tools can assist in identifying or detecting abuses.

Despite these innovations, a common limitation across these studies is their retrospective nature and lack of predictive capability. Most focused on classifying past violations or grouping similar cases but did not attempt to forecast when and where future violations might occur. Additionally, few studies addressed the integration of both statistical and deep learning models in a hybrid framework capable of capturing both seasonal trends and sudden non-linear dynamics.

This study addresses these gaps by introducing a novel hybrid predictive modeling framework — the Predictive Human Rights Violations Model (PHRVM) — which integrates SARIMA and RNNs. SARIMA enables modeling of structured seasonality (e.g., election-related repression), while RNN captures unexpected spikes linked to political unrest or security crackdowns. By combining these techniques, the study provides a more comprehensive and accurate forecasting tool tailored to civic space dynamics in Eastern Africa.

In doing so, this research contributes to advancing data science approaches in human rights monitoring in several key ways:

1. It moves beyond classification and descriptive analysis into time-aware forecasting.
2. It adapts forecasting models traditionally used in other fields like finance to the social and civic domain.
3. It demonstrates how narrative, qualitative data from reports can be structured and transformed for predictive analytics.
4. It offers a replicable framework that can be adapted for other regions or forms of social unrest.

Table 1: Summary of Studies Using Data Science in Human Rights Research

Study	Methodology	Dataset	Findings	Description
Medvedeva et al. (2020)	SVM, NLP	ECtHR cases	Predicted legal decisions (79% accuracy)	Highlights machine learning use in human rights but lacks predictive analysis
Marazzi (2016)	–	Human rights reports (2010–2015)	Developed automated data pipeline	Supports data-driven human rights research but lacks predictive modeling
Ran et al. (2023)	NLP, Decision Trees, Random Forest	HRD attack reports	85% accuracy in classifying attacks	Identifies key HRD risk factors but lacks predictive modeling
Smith (2020)	K-Means, LCA, Geospatial Analysis	CIRIGHTS dataset	Grouped African countries by violation patterns	Supports clustering but lacks forecasting models
Amnesty & HRW (2023–2024)	Qualitative analysis	Global human rights reports	Identified trends in HRD violations	Highlights need for predictive analytics
Quinn et al. (2018)	Remote sensing, ML, Supervised Classification	Satellite imagery (NASA, Google Earth)	Detected destruction as proxy for violations	Demonstrates AI-driven monitoring; could supplement ground-based data
Alhelbawy et al. (2020)	NLP, Crowdsourcing, Sentiment Analysis	Social media (Twitter), Crowdsourced data	Real-time human rights monitoring via Ceasefire Iraq	Supports real-time monitoring; future work could integrate NLP into forecasting

Chapter 3

Methodology

3.1 Philosophical Paradigm

In this study, we adopted an interpretivist philosophical paradigm to delve into the complex landscape of human rights violations against activists in East and Horn of Africa. By using this perspective, we intended to learn more about the fundamental causes and current patterns that drive human rights violations. Our exploration begun with a dataset spanning from 2020 to 2023, compiled violations from documented reports by reputable organizations such as Al Jazeera Al Jazeera, n.d., Cable News Network (CNN) CNN, n.d., Amnesty International Amnesty International, 2024, Crisis Watch Crisis Group, n.d., Addis Standard Addis Standard, 2024, Missing Voices Missing Voices, 2024, Article 19, Front Line Defenders, Armed Conflict Location and Event Data Project (ACLED) ACLED, 2024, Sudan Tribune Sudan Tribune, n.d., British Broadcasting Corporation (BBC) BBC, n.d., Reporters Without Borders, Committee to Protect Journalists (CPJ), Africa News Africa News, n.d., and other reliable sources.

While the original data consisted of qualitative narratives comprising incident descriptions, threats, and event contexts, it was transformed into structured, categorical formats suitable for statistical and machine learning analyses. Each incident was encoded with a set of predefined variables (e.g., type of violation, country, group targeted, and date), allowing for predictive analytics on what was once narrative-based data.

In the first objective, we plan to identify current patterns of human rights abuses in Eastern Africa with respect to their frequency, nature, and geographic distribution. We aim to obtain strategic insights for advocacy initiatives that will improve the protection of HRDs and overall human rights observance.

A crucial aspect of the study consists the development of a predictive model to assess human rights violations against HRDs based on patterns from historical data for 2020 through 2023. With the interpretivist paradigm, predictive techniques that could capture the complicated, context-sensitive nature of repression in Eastern Africa will be identified.

To ensure the reliability and applicability of our predictive model, the third objective of this study focuses on evaluating its accuracy and performance in forecasting human rights violations using appropriate validation metrics. This aligns with the interpretivist philosophical paradigm, as it emphasizes understanding the contextual and dynamic nature of human rights violations rather than treating them as purely statistical occurrences. By assessing the model's effectiveness, we aim to determine whether it meaningfully captures historical patterns and provides reliable forecasts that can inform early intervention strategies.

The evaluation process involved quantitative validation metrics, such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), to measure the deviation between predicted and actual violations in 2024. Additionally, hypothesis testing using a paired t-test was conducted to establish whether the PHRVM significantly improves forecasting accuracy over a naïve baseline model. The findings indicated that the PHRVM effectively balances short-term fluctuations with long-term trends, reinforcing its practical applicability for human rights monitoring organizations.

3.2 Study Site

The study encompasses a comprehensive examination of human rights violations against HRDs across 11 countries in East and Horn of Africa. These countries include Djibouti, Eritrea, Ethiopia, Kenya, Burundi, Tanzania, Somalia & Somaliland, Rwanda, South Sudan, Sudan and Uganda.These locations provide a comprehensive representation of the human rights landscape across the East and Horn of Africa region, enabling an understanding of the challenges and violations faced by HRDs in different contexts.

3.3 Research Design

The study employed a quantitative research design with qualitative data input. Primary analysis was statistical and predictive in nature, while raw data consisted of qualitative narratives including descriptions of incidents, threatened types, and contextual details. These data were systematically organized and categorized into quantitative variables for modeling. The study also adopted a cross-sectional design for analyzing the human rights violations against Human Rights Defenders (HRDs) in Eastern Africa between 2020 and 2023.

This design was appropriate for this research because it allowed for the collection and analysis of data across multiple variables simultaneously, capturing both descriptive and inferential insights into human rights abuses in the region. Cross-sectional studies are regarded as observational designs that study data from a population at any given point in time. This approach is widely used in social sciences and human rights research when the goal is to identify correlations, trends, and risk factors within a defined time frame Bryman (2012). Since the study sought to analyze both historical trends and develop a forecasting model, a cross-sectional framework was the best fit, providing a solid foundation for statistical and machine learning-based modeling.

To ensure a comprehensive representation of human rights violations, data was collected from reputable sources, including human rights organizations, global news agencies, and verified open-access datasets. The dataset included a wide range of incidents across different countries in the region, capturing variations in violation types, frequency, and geographic distribution. Unlike longitudinal studies, which track changes over time in a controlled setting, the cross-sectional approach enabled this research to identify key patterns and relationships in the data while maintaining computational efficiency.

In the analytical phase, a combination of descriptive statistics and machine learning algorithms was employed to extract meaningful insights from the dataset. Descriptive statistics were used to summarize the data, highlight distribution patterns, and identify key trends, a methodology supported by Sampaio (2024), who emphasizes that descriptive analysis helps convert raw data into interpretable insights for decision-making. By applying visualization techniques, such as bar charts, heatmaps, and geospatial mapping, the study uncovered which regions experienced the highest violations and the key risk factors contributing to these abuses.

Python programming was used alongside relevant data science libraries such as pandas, scikit-learn, and TensorFlow for data preprocessing, statistical analysis, and the development of a predictive model. To achieve this, SARIMA (for modeling seasonal trends), RNN (for short-term fluctuations) and the PHRVM was evaluated based on its ability to capture temporal trends, handle seasonality, and predict violations accurately. Through a structured comparative analysis, the PHRVM emerged as the optimal forecasting approach, as it successfully balanced short-term fluctuations (RNN) with structured seasonal patterns (SARIMA).The selection of PHRVM was based on its performance in quantitative evaluation metrics, such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

The final objective focused on evaluating the accuracy and reliability of the chosen model, ensuring that its forecasts were statistically valid and practically useful. The model's predictions for 2024 violations were compared against actual observed violations, and its performance was assessed using MAE and RMSE metrics. Additionally, hypothesis testing through a paired t-test was conducted to determine whether the PHRVM significantly outperformed a naïve baseline forecast.

3.4 Area of Study

The study encompasses a comprehensive examination of human rights violations against HRDs across 10 countries in Eastern Africa. These countries include Djibouti, Eritrea, Ethiopia, Kenya, Burundi, Tanzania, Somalia & Somaliland, Rwanda, South Sudan and Uganda.These locations provide a comprehensive representation of the human rights landscape across the Eastern Africa region, enabling an understanding of the challenges and violations faced by HRDs in different contexts.

3.5 Study Population

3.5.1 The Inclusion Criteria

The study population for this research consists of civic violations on HRDs in Eastern Africa.It encompassed individuals or groups actively engaged in advocating for human rights in Eastern Africa, including those promoting freedom of expression, association, and peaceful assembly. Additionally, HRDs who have been documented as victims of human rights violations, as reported on known sites or news channels, are within the scope of our study.

3.5.2 The Exclusion Criteria

Exclusion criteria involve HRDs who do not meet the specified inclusion criteria. This delineation ensures a focused investigation into the challenges and experiences faced by HRDs in the Eastern Africa, particularly those documented as victims of rights violations.

3.6 Source of Information

This research relied on secondary data sources from reputable human rights organizations, international news agencies, and conflict monitoring databases. Geopolitical assessments, comprehensive event reports, and regional trends in civic space restrictions were all provided by these sources. The statistical analysis, geographic mapping of violations, and predictive modelling employed were all based on the data gathered from these sources.

First, we collected data from human rights organisations that includes Global and regional human rights organisations, which frequently release reports on civic space, freedom of expression, and attacks on HRDs. These resources included incident-level information on violations, governmental actions, and patterns of repression of public space. These include.

Secondly, we used Conflict and Crisis Monitoring Databases to track violence, repression trends, and civic unrest, this study used open-access conflict monitoring platforms, which pro-

Table 2: Organizations and Platforms Used for Data Collection

Name	**Description**	**Link**
Amnesty International	Reports on human rights violations, political repression, and threats to HRDs worldwide	https://www.amnesty.org
Article 19	Focuses on freedom of expression and digital censorship	https://www.article19.org
Front Line Defenders	Provides case reports on individual HRDs facing persecution	https://www.frontlinedefenders.org
Committee to Protect Journalists (CPJ)	Tracks press freedom violations, journalist arrests, and media suppression	https://www.cpj.org
Reporters Without Borders (RSF)	Monitors global press freedom and documents abuses against journalists	https://www.rsf.org
DefendDefenders	Focuses on human rights violations against HRDs in Eastern Africa	https://www.defenddefenders.org

vided structured datasets on protests, arrests, government crackdowns, and security-related incidents as seen below.

Table 3: Human Rights Databases Used for Data Collection

Name	**Description**	**Link**
Armed Conflict Location & Event Data Project (ACLED)	An extensive database tracking political violence and protests globally	https://www.acleddata.com
Crisis Watch	A global early-warning platform for political conflicts and human rights issues	https://www.crisisgroup.org
Missing Voices Kenya	A civic space watchdog focusing on extrajudicial killings and police violence in Kenya	https://www.missingvoices.or.ke

3.7 Variables definitions

The dependent variable in this study is the Right Violated, which identifies the specific human rights that were breached in each recorded incident. This variable serves as the primary outcome of interest, categorizing violations such as freedom of expression, freedom of assembly, and the right to speech. Since each violation falls into a distinct category, it is measured at the nominal level and is essential for classification and predictive modeling to analyze patterns of repression and forecast future risks.

Several independent variables provide explanatory factors influencing human rights violations. The Category of HRD classifies defenders based on their roles, such as journalists, lawyers, activists, and opposition figures. Different HRD groups face varying levels of risk depending on their involvement in advocacy, media, or political activities, making this variable crucial in identifying trends in targeted repression. As a categorical classification, it is measured at the nominal level. Similarly, the Country variable identifies the geographic location of each violation. This variable plays a crucial role in spatial analysis and is also measured at the nominal level.

The Threats variable describes the nature of dangers faced by HRDs, including arbitrary arrests, judicial harassment, surveillance, and death threats. Understanding these threats helps categorize patterns of repression and identify emerging risks. This variable is also measured at the nominal level, as each threat falls into a predefined category. The Date of Incidence records when each violation occurred, making time an essential factor in trend analysis and forecasting. This variable is integral to the study's time-series modeling, particularly in SARIMA and RNN models, where temporal dependencies are key to predicting future violations. Since it represents an ordered sequence without a true zero point, it is measured at the interval level.

To give contextual understanding, a short Description variable provides a detailed narrative of each incident, including circumstances and background information. This qualitative data allows for a deeper analysis of human rights violations but remains nominal in measurement as it consists of text-based descriptions. The Source variable identifies where the data originates, including NGO reports, news articles, and legal documents. Given that sources differ in credibility and scope, this variable is important for validation and credibility assessment. Like other categorical variables, it is measured at the nominal level.

Finally, Seasonality captures the periodic nature of human rights violations, identifying whether incidents occur during elections, protests, or policy changes. Given that human rights violations often escalate during politically sensitive periods, seasonality serves as an important predictor in forecasting models.

Table 4: Variables Included in the Human Rights Dataset

Variable	**Description**
Category of HRD	Categorizes the HRD based on their role or engagement in human rights activities.
Country	Identifies the country where the incident occurred.
Threats	Describes the nature of threats faced by the HRD.
Right Violated (Target Variable)	Identifies the specific human rights that were violated in the incident.
Date of Incidence	Specifies the date of the incident.
Full Description	Provides a detailed account of the incident, including relevant circumstances.
Source	Indicates the sources of information about the incident.
Seasonality	Events when the violation occurred.

3.8 Quality/Error Control

To ensure a high level of reliability for the study, data were sourced from only credible international organizations extending to Amnesty International, Human Rights Watch, Reporters Without Borders, ACLED, and DefendDefenders. These organizations respect strict rules of fact-checking and verification which prove that the recorded violations are credible and evidence-based. News agencies such as the BBC, CNN, and Al Jazeera were also incorporated, as they report real-time human rights violations. Verifying each human rights violation against several independent sources guaranteed that the dataset was comprised of verified and integrated information minimizing the possibility of misinformation or bias. When collecting the data, thorough documentation of data sources, addressing conflicts through careful review, and cross referencing information from several sources was done to ensure correctness.

During data cleaning verification was performed to check for duplicates, inconsistencies, and missing data, ensuring that each incident recorded in the dataset accurately reflected documented events.Furthermore, precautions were made to protect the confidentiality and privacy of those named in the data. Personal identifying information was anonymized or removed where necessary to protect the identities of victims, witnesses, and other involved parties. The use of standardized classification criteria for violation types, affected HRDs, and repression tactics further enhanced data validity by preventing ambiguous or overlapping categorizations.

Predictive modeling was done with SARIMA and RNN. Therefore, the models were validated using Mean Absolute Error and Root Mean Squared Error, which give a measure of deviation from the actual violation trends. The results were assessed, and hyperparameters were tuned for optimal model performance. Data were split into training (80%) and test (20%) subsets, and evaluation was performed on unseen data prior to real-world forecasting. Geographic stratification was used to ensure that countries with lower media freedom (e.g., Eritrea, Djibouti) were not underrepresented in the analysis.

3.9 Data processing and analysis

The data processing and analysis phase was structured to address the research questions by extracting meaningful insights from the dataset and ensuring the reliability of the predictive model. The analysis was conducted in four key stages: descriptive analysis, clustering, predictive modeling, and model evaluation.

3.9.1 Descriptive Analysis

First, descriptive statistical methods were used to identify patterns, trends, and key risk factors associated with human rights violations against HRDs. Exploratory data analysis (EDA) techniques, such as visualizing frequency distributions, geographic clustering, and seasonality analysis, helped uncover the spatial and temporal dynamics of violations in Eastern Africa. This stage established a foundational understanding of where, when, and how violations occur, providing key insights for the next phases.

3.9.2 Clustering Analysis with K-Means

To further investigate patterns in human rights violations, K-Means clustering was applied to group countries based on similarities in the nature and frequency of violations. The dataset was first standardized using feature scaling (StandardScaler) to ensure that differences in scales between variables did not bias the clustering algorithm. The Elbow Method was used to determine the optimal number of clusters, ensuring that the segmentation accurately reflected meaningful groupings. The clustering process revealed distinct regional patterns, identifying high-risk zones and areas with emerging threats.

This clustering approach provided critical insights into how different countries experience violations, allowing for more targeted forecasting in the next stage. By identifying regions with similar violation patterns, the study improved its predictive modeling accuracy, ensuring that country-specific trends were accounted for in forecasting human rights violations.

3.9.3 Predictive Modeling

Following the clustering analysis, predictive modeling techniques were applied to develop a forecasting framework for human rights violations. Multiple models, including SARIMA, LSTM, RNN, SARIMA and LSTM, and SARIMA and RNN were tested to assess their ability to capture trends and fluctuations in violations over time. A structured performance evaluation was conducted, and the PHRVM model was developed as the optimal forecasting model. This model successfully combined short-term adaptability (RNN) with long-term seasonality modeling (SARIMA), providing a balanced and robust approach to forecasting violations.

3.9.4 Model Evaluation and Hypothesis Testing

The final objective of the study focused on evaluating the accuracy and reliability of the predictive model. The model's forecasts for 2024 violations were compared against actual observed violations, with Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) serving as key performance indicators. Additionally, hypothesis testing was conducted using a paired t-test to determine whether the PHRVM model significantly outperformed a naïve baseline forecast.

The results confirmed that the PHRVM model significantly improved forecasting accuracy, making it a practical and effective tool for early warning systems and risk assessment in human rights monitoring. By combining statistical, machine learning, and clustering techniques, this study ensures that predictions are not only data-driven but also contextually relevant, enhancing the ability of human rights organizations to anticipate and mitigate risks against HRDs.

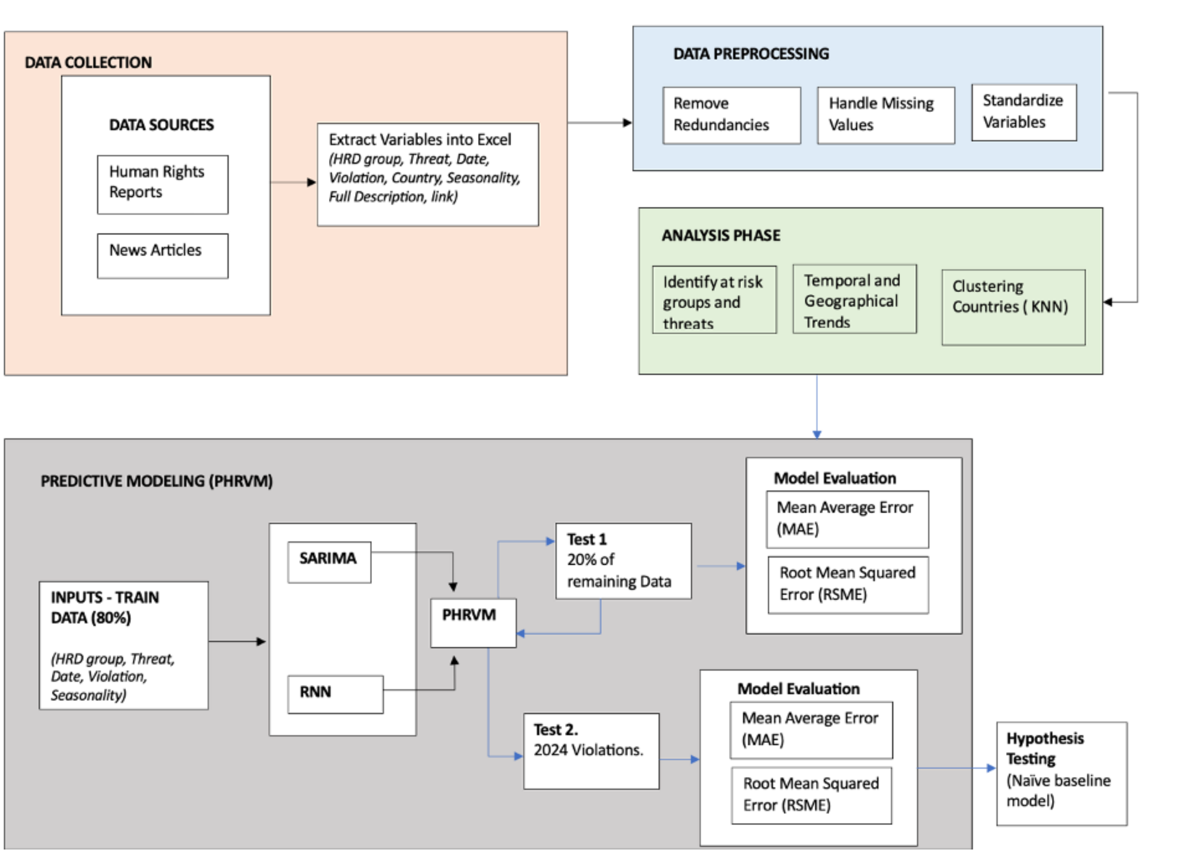

Figure 1: Conceptual and analytical framework for PHRVM

3.10 Ethical considerations

The dataset contained information about arrests, enforced disappearances, censorship, and persecution of HRDs, which could place individuals at risk if improperly disclosed. Since the data was collected from public reports and human rights organizations, no personal identifiers (such as names or direct contact details of HRDs) were included in the dataset. Any references to specific cases were anonymized, ensuring that information was used only for statistical and analytical purposes, not for identifying individuals.

All sources used in this study were from verified, reputable organizations, such as Amnesty International, Human Rights Watch, ACLED, and Reporters Without Borders, ensuring that only fact-checked and evidence-based reports were included. Additionally, the research followed strict data validation protocols, cross-referencing multiple independent sources to avoid misinformation and biases. The study avoided exaggeration, ensuring that findings were reported objectively and accurately.Links to the violations were also mandatory to ensure the information is valid.

Since the research relied on existing public data rather than primary interviews or surveys, formal consent from participants was not required. However, ethical considerations still applied to the use of third-party reports. This study followed the terms and conditions of data providers, ensuring compliance with usage policies of human rights organizations, open-source databases, and news agencies. Where required, attribution was provided to the original sources, recognizing the work of organizations that document human rights violations.

Because HRDs work in environments hostile to state surveillance and repression, special precautions were taken to prevent possible misuse of the research. The precise locations of activists and the sensitive operational details of human rights organizations were not disclosed by the research making the research null for possible tracking or targeting HRDs.

Chapter 4

Data Analysis, Presentation and Interpretation of Findings

4.1 Data Cleaning and Preprocessing

Data cleaning is a crucial step in ensuring the integrity and reliability of any analytical process, where inconsistencies in reporting and missing values can distort findings. The dataset used in this study contains human rights violations recorded in the Eastern Africa from 2020 to 2023. Given that the data is sourced from multiple reports, including NGOs, news articles, and government documents, inconsistencies in formatting, missing information, and variations in terminology are common. These inconsistencies can affect statistical analysis and machine learning predictions, making a thorough data cleaning process essential. The preprocessing steps described below focus on handling missing values, standardizing categorical data, and ensuring consistency across key fields such as country names, violation types, and affected groups.

4.1.1 Handling Missing Values

Instead of removing records with missing values, which could lead to the loss of critical information, missing values were replaced with the term "Not Applicable". This approach preserves the dataset's completeness while ensuring that null values do not interfere with categorical data processing and classification models. By replacing missing values, the dataset remains structurally intact without compromising analytical accuracy. This is particularly useful for later machine learning tasks, where missing values could cause models to fail or introduce bias.

4.1.2 Standardizing Categories

In the dataset used for this study, inconsistencies were observed in country names, groups affected by violations, and types of human rights violations. Without standardization, the same entity might be recorded under different labels, leading to fragmentation in statistical analysis. For example, the country "South Sudan" appeared in multiple formats such as "south Sudan" and "South Sudan", which were unified under "SouthSudan" to maintain consistency across records. Similarly, "Somaliland" was merged into "Somalia" to prevent duplicate geographic entries that could distort spatial trend analysis. Another major issue in the dataset was the variation in how affected groups were labeled. The term "journalists" appeared in multiple forms such as "journalist," "Journalists," and "journalists", which were all standardized to "Journalist." Likewise, opposition groups had inconsistent entries such as "opposition," "Opposition leaders," and "Opposition politicians," which were all consolidated under "Opposition" to allow for proper categorization. The "Type of Violation" column also contained discrepancies, with multiple variations of "Freedom of Expression" recorded as "Freedom of expresssion," "freedom of expression," and "Access to Information", all of which were standardized to "Expression."

4.2 Analyze patterns of human rights violations against HRDs in the Eastern Africa, including the key threats, targeted groups, and trends over time.

Understanding the patterns of human rights violations against HRDs in Eastern Africa is critical for assessing the scale, key threats, targeted groups, and trends over time. This section presents an analysis of the frequency of violations per country, the distribution of key threats, and regional trends based on the dataset. Using data visualization techniques such as bar charts, count plots, and pie charts, the findings provide a comprehensive view of the violations occurring in different regions. The results, as visualized in the attached figures, were obtained through Python-based data analysis and visualization using the Seaborn and Matplotlib libraries.

4.2.1 Distribution of Violation Types by Country

To assess the prevalence of different types of human rights violations across Eastern Africa, a pivot table was created to count the occurrences of each "Type of Violation" by "Country". This allows for a structured view of how violations are distributed regionally. The results (as shown in figure 1) indicate that Freedom of Expression violations are the most frequently recorded type of violation across multiple countries, with Somalia (89), Uganda (71), and Ethiopia (55) reporting the highest numbers. Violations related to Freedom of Association appear most frequently in Uganda (40) and Tanzania (21), while Peaceful Assembly violations are most frequent in Kenya (33) and Uganda (22). The General category of human rights violations is present in all countries but with lower frequencies.

South Sudan and Eritrea show lower absolute numbers compared to other nations in the dataset. However, the overall distribution varies across different categories, with some countries exhibiting higher cases in specific types of violations compared to others. The dataset does not suggest uniformity across the region, indicating variations in the number and type of reported violations across different locations.

Type of violation	Association	Expression	General	Peaceful Assembly
Country				
Burundi	14	13	2	0
Djibouti	2	5	0	1
Eritrea	10	8	0	0
Ethiopia	17	56	4	11
Kenya	18	36	5	33
Not Applicable	0	2	0	0
Rwanda	6	26	6	0
Somalia	2	89	1	3
SouthSudan	5	25	4	5
Sudan	2	13	2	14
Tanzania	21	30	1	8
Uganda	40	71	2	22

Figure 2: Pivot table showing the amount of civic violations across the region

4.2.2 Geographic Distribution of Human Rights Violations

The bar chart visualization in figure 3 presents the frequency of human rights violations by country, showing variations in the number of reported cases across different nations in Eastern Africa. The data indicates that Uganda, Ethiopia, and Somalia recorded the highest number of violations, with Uganda reporting the most cases. Ethiopia follows as the second highest, while Somalia ranks third in reported violations.

The dataset also shows that Tanzania, Kenya, and Burundi have a notable number of recorded violations, with reported cases spanning different categories of human rights abuses. In contrast, Djibouti, Eritrea, and Rwanda have lower recorded violation counts in the dataset. However, there are variations in the types of violations across countries, as further examined in the following sections.

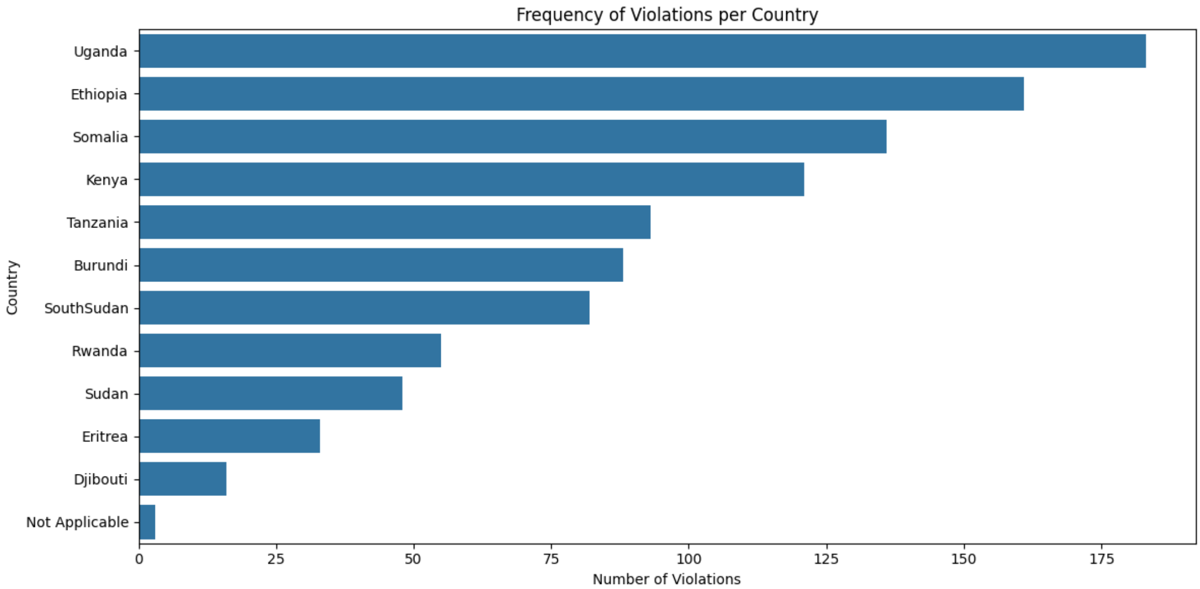

Figure 3: Horizontal bar chart showing the frequency of violations across the region

4.2.3 Key threats against human rights defenders

To analyze the nature of threats faced by HRDs, the column "Threats" was analysed and a pie chart was generated to display the top 10 most common threats recorded in the dataset. The pie chart as seen in figure 3 indicates that arbitrary detention (27.4%) and arbitrary arrests (23.7%) are the most common threats faced by HRDs, collectively making up over 50% of recorded violations. These forms of repression are widely used by state actors to silence journalists, activists, and opposition members. HRDs advocating for democracy, political reforms, and press freedom are frequently arrested under vague national security laws, terrorism charges, or sedition accusations.

Other significant threats include attacks (9.0%), killings (7.1%), and violent repression (7.1%), reflecting the high-risk environment faced by HRDs in conflict-prone regions such as Somalia, Ethiopia, and South Sudan. Censorship, police brutality, and judicial harassment also appear prominently, highlighting the legal and institutional challenges that HRDs encounter in restrictive civic spaces.

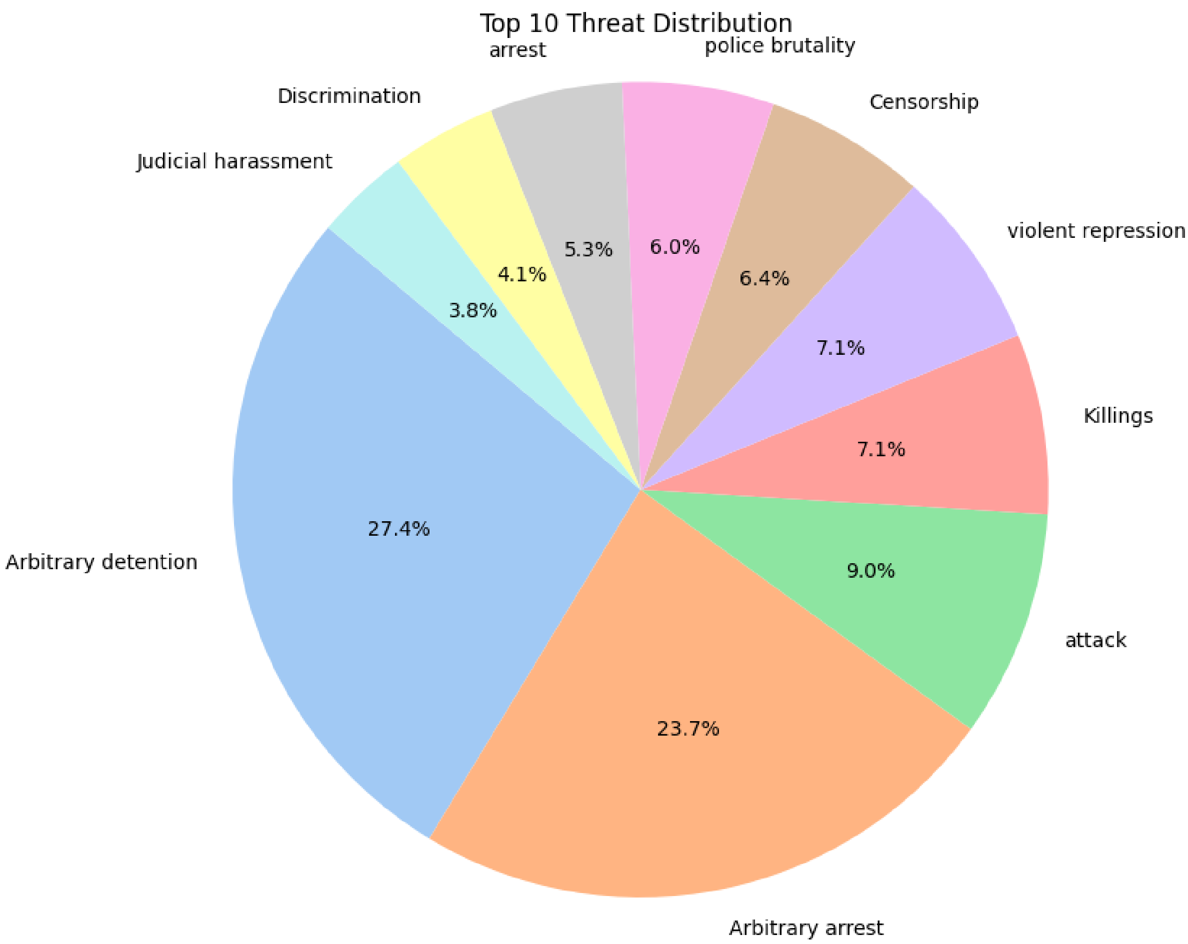

Figure 4: Pie chart showing the top 10 distribution of threats from the period 2020-2023

4.2.4 Human rights violations

A detailed analysis of human rights violations against Human Rights Defenders (HRDs) in Eastern Africa for the years 2020 - 2023 was conducted using multiple visualizations, including bar charts and a choropleth map. These visual representations provide insights into the most frequently violated rights, the groups most affected, the nature of threats HRDs faced, and the geographic distribution of reported cases.

4.2.4.1 Human rights violations in 2020

Distribution of Rights Violated

Figure 5 presents the frequency of different types of rights violations recorded in 2020. The dataset shows that violations related to freedom of expression were the most prevalent, followed by restrictions on peaceful assembly and freedom of association. The higher frequency of expression-related violations suggests that restrictions on speech, censorship, and media suppression were among the most commonly reported forms of repression in this period.

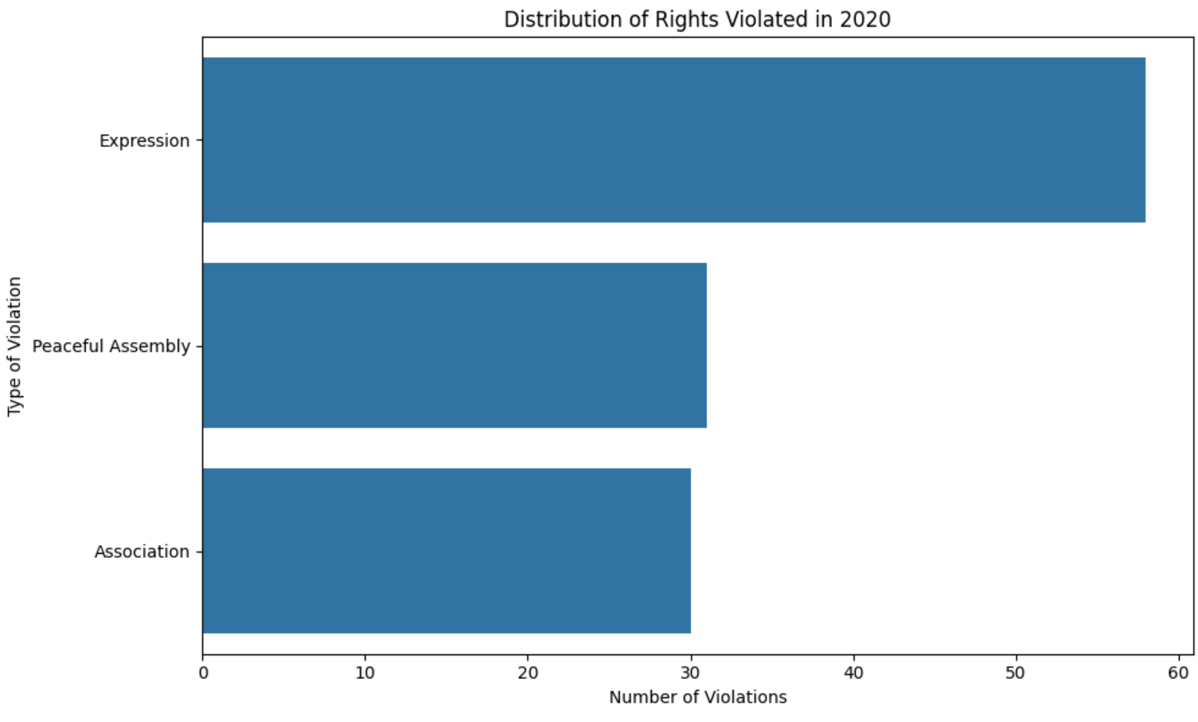

Figure 5: Bar chart showing the distribution civic violations in 2020

Groups Most Affected

Figure 6 illustrates the top 15 groups targeted in human rights violations in 2020. The dataset shows that journalists were the most frequently affected, followed by opposition members and protesters. Other affected groups included minorities, media personnel,HRDs, teachers, civilians, politicians, and religious minorities. The distribution highlights the varying degrees of repression experienced by different groups, with journalists and opposition figures facing a disproportionately high number of violations.

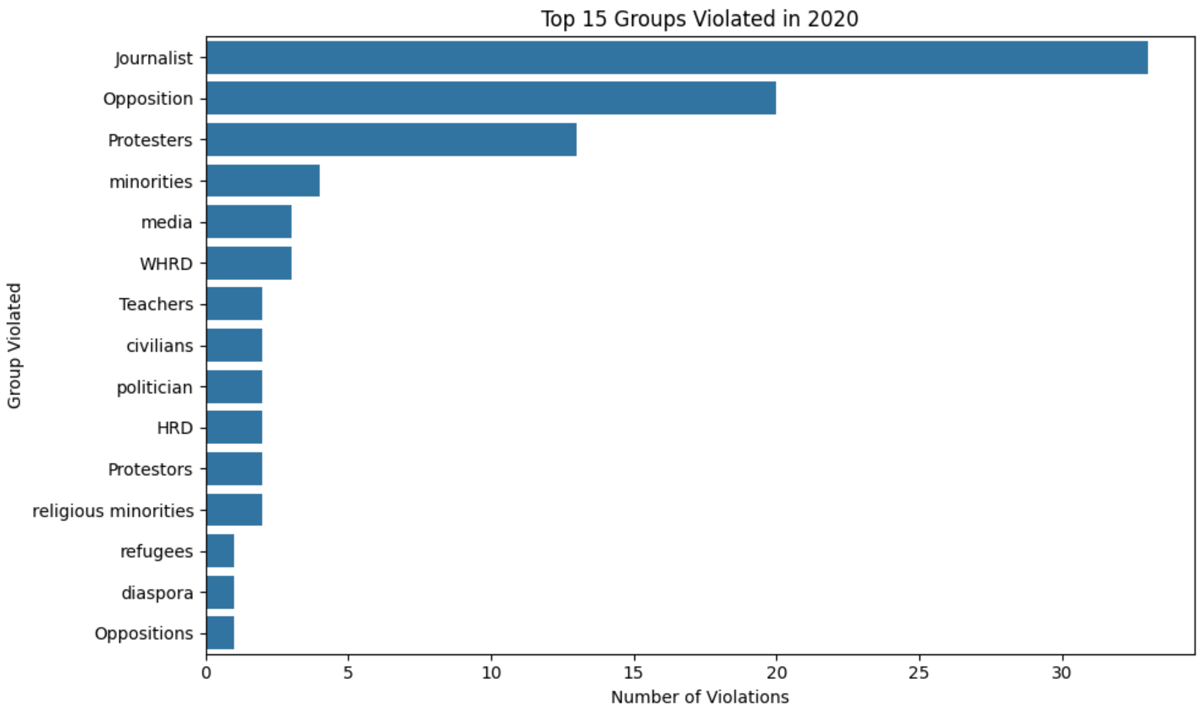

Figure 6: Horizontal bar chart showing the top 15 groups violated in 2020

Nature of Threats Against HRDs

Figure 7 provides an overview of the most commonly reported threats to HRDs in 2020. The dataset indicates that attacks, arrests, and arbitrary arrests were the most frequently documented threats. Other reported threats included censorship, arbitrary detention, violent repression, police brutality, enforced abductions, and killings. These findings reflect the diverse range of threats HRDs faced, indicating a widespread use of both legal and extrajudicial measures against activists, journalists, and opposition figures.

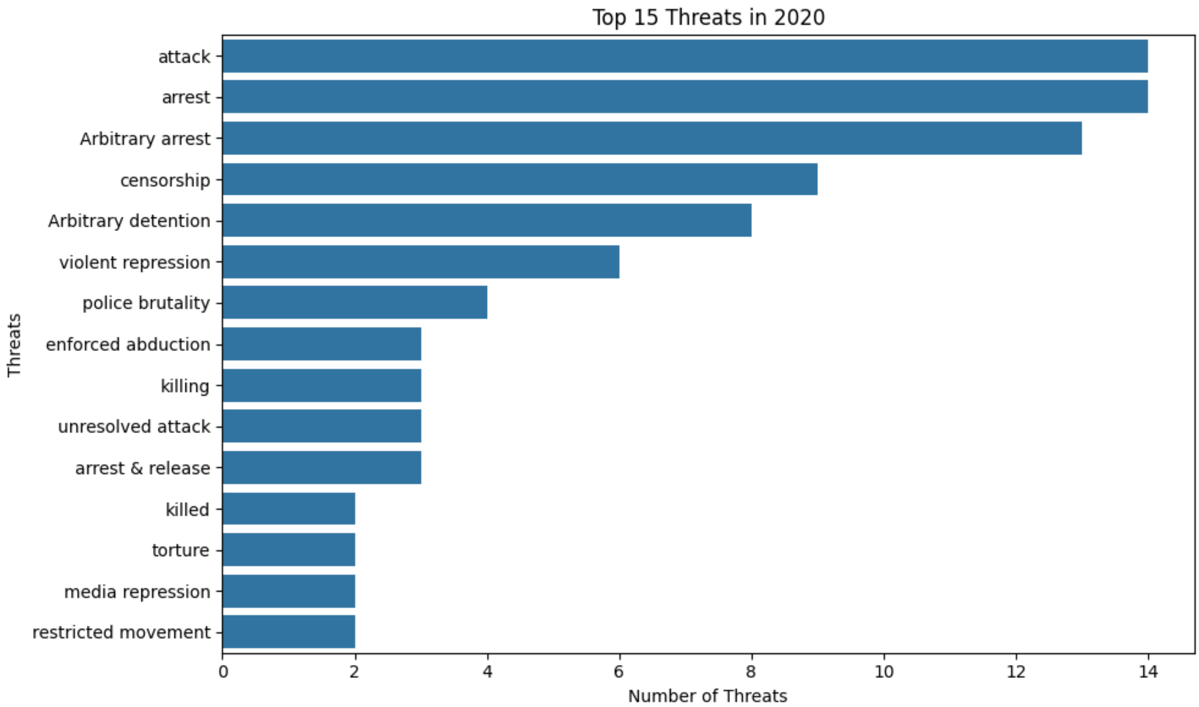

Figure 7: Horizontal bar chart showing top 15 threats faced by HRDs in 2020

Geographic Distribution of Violations

The choropleth map visualizes the distribution of recorded human rights violations across Eastern Africa in 2020. The dataset reveals significant variations in the number of violations across different countries. The intensity of reported violations, represented by darker shades, indicates that certain nations experienced higher frequencies of documented human rights abuses compared to others. The map was created using Plotly, which enables the generation of interactive visualizations. The geographic data was processed using GeoPandas to map the locations of human rights violations to specific countries.

The dataset as seen in figure 8 reveals that certain countries had higher reported frequencies of violations compared to others on the map.The results from the visualization indicate the following:

- Uganda, Ethiopia, and Somalia recorded the highest number of human rights violations against HRDs in 2020. These countries show the darkest shades on the map, indicating a higher intensity of recorded cases compared to other nations in the region.

- Kenya, Tanzania, and Burundi also reported a notable number of violations, though at a slightly lower frequency than the top three countries.
- South Sudan and Rwanda exhibit moderate levels of recorded violations, with fewer reported cases than their neighboring countries.
- Djibouti and Eritrea show the lowest number of documented violations, appearing in the lightest shades on the map.

4.2.4.2 Human rights violations in 2021

Distribution of Rights Violated

The analysis as seen from the bar chart in figure 8 indicates that violations related to freedom of expression were the most frequently reported, followed by restrictions on freedom of association and peaceful assembly. The number of expression-related violations increased compared to 2020, suggesting that restrictions on speech, censorship, and press freedom remained significant concerns during this period. Violations of peaceful assembly were the least frequently recorded among the three categories.

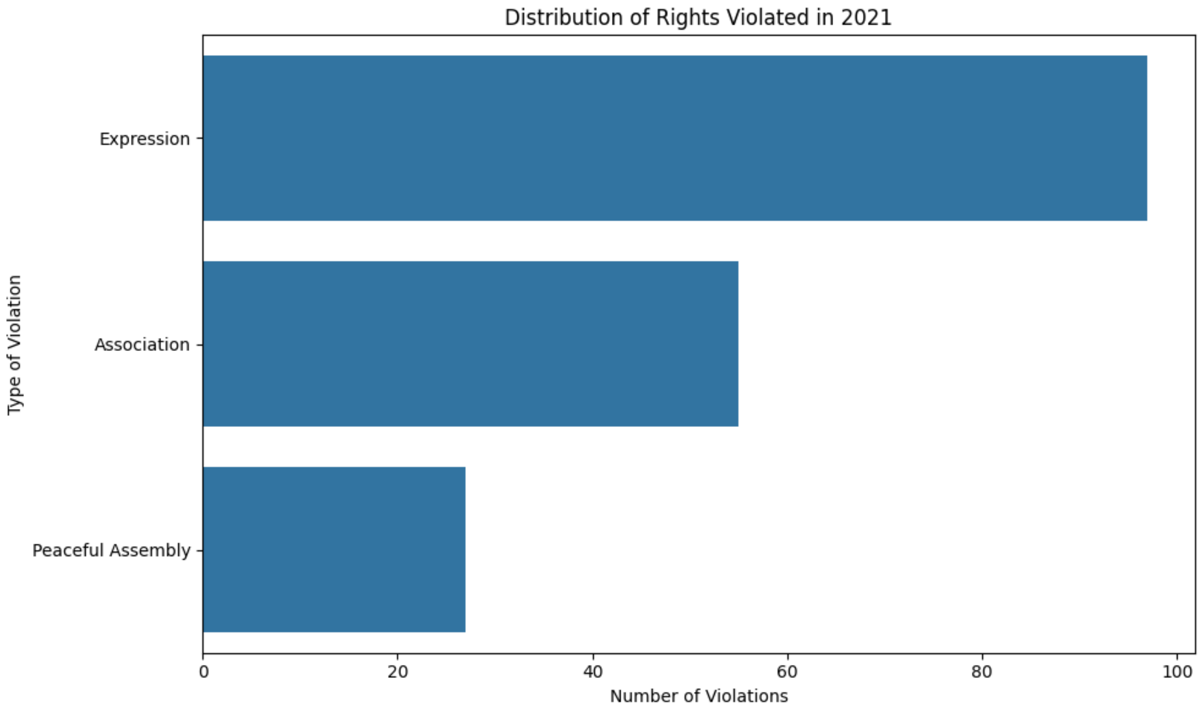

Figure 8: Bar chart showing the distribution civic violations in 2021

Groups Most Affected

In 2022, Journalists were the most frequently targeted group, with significantly more violations recorded than any other category. Opposition members and protesters were the second and third most frequently affected groups. Other affected groups included activists, citizens, Christians, political opposition figures, HRDs, media outlets, civilians, and minorities. The data indicates that a diverse range of individuals and organizations were subjected to repression, censorship, and targeted attacks.

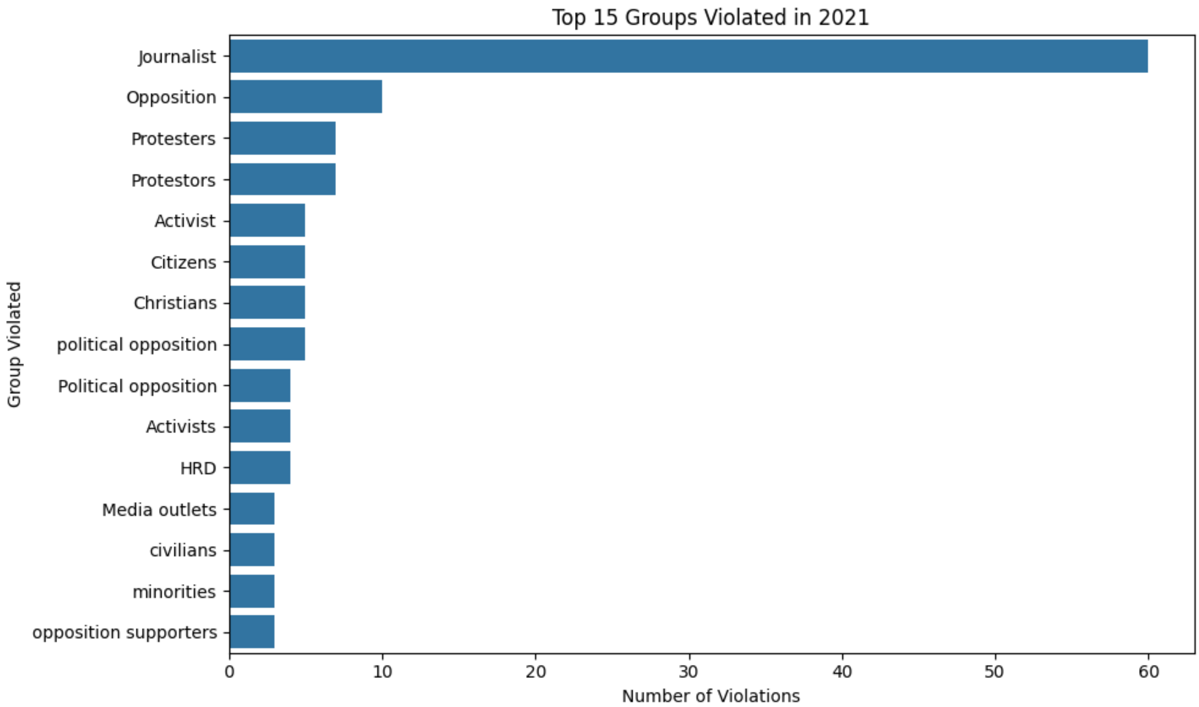

Figure 9: Horizontal bar chart showing the top 15 groups violated in 2021

Nature of Threats Against HRDs

Figure 10, shows that the most common threats were arbitrary arrest, arbitrary detention, violent repression, and police brutality. Other frequently reported threats included attacks, internal conflict, imprisonment, enforced abduction, and media repression. The analysis shows an increase in the use of arbitrary legal measures such as detention and imprisonment, along with physical violence in various forms, compared to the previous year.

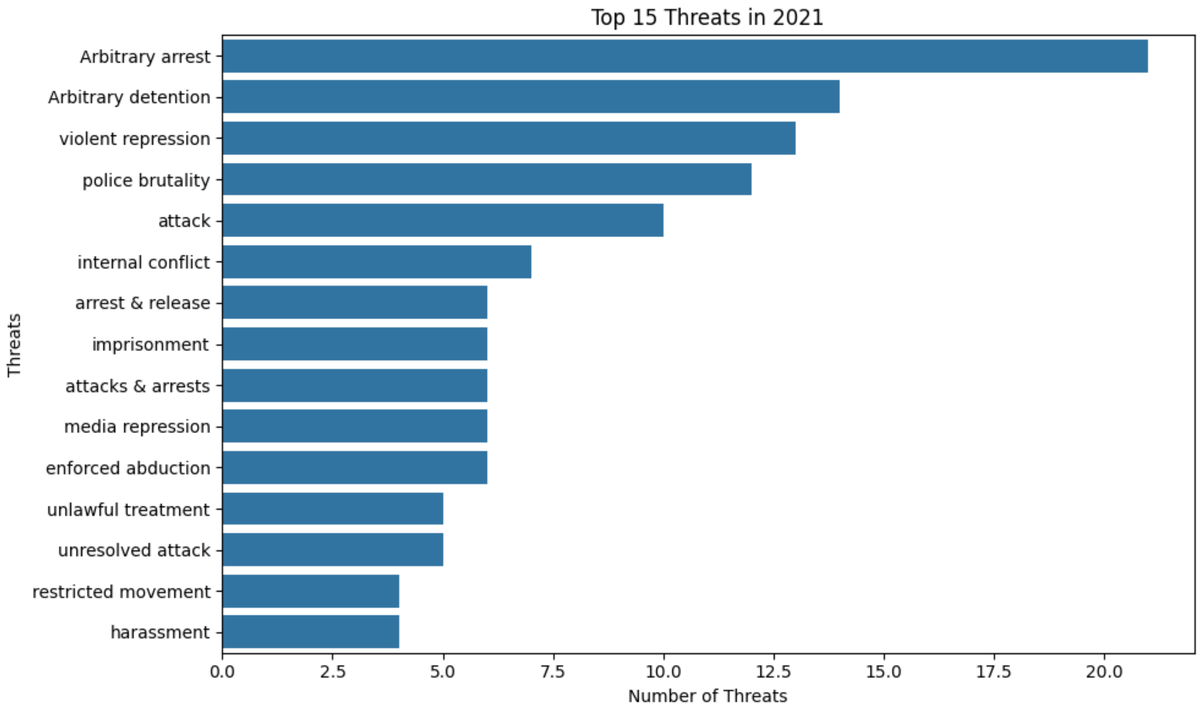

Figure 10: Horizontal bar chart showing the top 15 threats violated in 2021

Geographic Distribution of Violations

The choropleth map below shows that certain countries recorded a higher frequency of violations compared to others. The results are as follows:

- Uganda, Ethiopia, and Somalia recorded the highest number of human rights violations, maintaining their positions as the most affected countries. These nations exhibit the darkest shades on the map, indicating a high concentration of reported cases.
- Kenya, Tanzania, and Sudan also reported significant numbers of violations, though slightly lower than the top three countries.
- South Sudan, Rwanda, and Burundi show moderate levels of recorded violations, with fewer reported cases than their neighboring countries.
- Djibouti and Eritrea recorded the lowest number of documented violations in the dataset, appearing in the lightest shades on the map.

Figure 11: Distribution of violations by country in 2021

4.2.4.3 Human rights violations in 2022

Distribution of Rights Violated In 2022, violations related to freedom of expression were the most frequently reported, followed by restrictions on peaceful assembly and freedom of association. The frequency of expression-related violations increased significantly compared to previous years, showing that restrictions on press freedom, speech, and censorship were widespread in the region. Violations of peaceful assembly and association were also present but at comparatively lower levels.

Figure 12: Bar chart showing the distribution civic violations in 2022

Groups Most Affected Figure 13 reveals that journalists were the most frequently targeted group, with a significantly higher number of recorded violations than any other category. Protesters were the second most affected group, followed by critics, minorities, activists, authors, and opposition members. Other affected groups included political opposition figures, civilians, media workers and minority groups. The data highlights the disproportionate targeting of journalists, alongside repression against protesters and opposition figures.

Figure 13: Horizontal bar chart showing the top 15 groups violated in 2022

Nature of Threats Against HRDs The bar chart below presents the top 15 threats recorded against HRDs in 2022. The dataset indicates that arbitrary arrest was the most common threat, followed by killings, arbitrary detention, restrictive laws, and censorship. Other frequently reported threats included violence against protesters, detention of journalists, physical assault, suppression of protests, enforced disappearances, intimidation, and judicial harassment. The data suggests that both legal and extrajudicial measures were widely used against HRDs, with an increase in physical violence and repression compared to previous years.

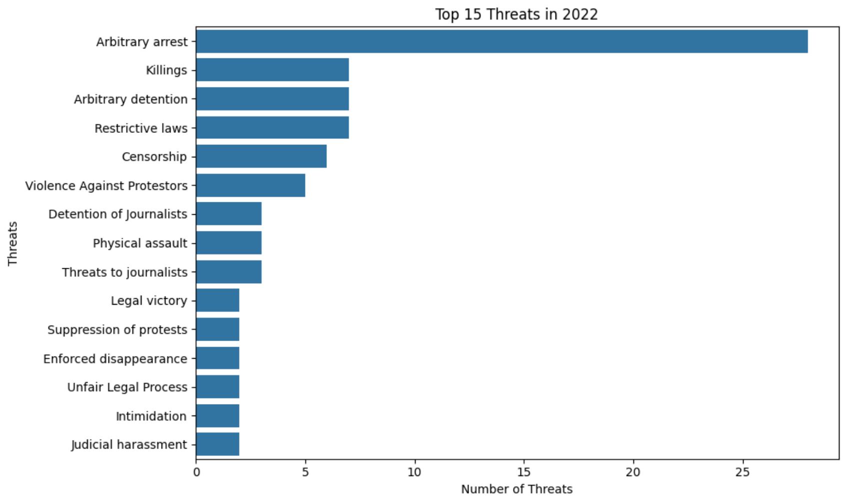

Figure 14: Horizontal bar chart showing top 15 threats faced by HRDs in 2022

Geographic Distribution of Violations The choropleth map for 2022 shows the following outcomes:

- Uganda, Ethiopia, and Somalia recorded the highest number of human rights violations, maintaining their positions as the most affected countries. These nations exhibit the darkest shades on the map, reflecting a high concentration of recorded cases.
- Kenya and Sudan also reported significant numbers of violations, although slightly lower than the top three countries.
- Tanzania, South Sudan, and Rwanda showed moderate levels of recorded violations, with fewer reported cases than their neighboring countries.

- Djibouti and Eritrea had the lowest number of documented violations, appearing in the lightest shades on the map.

Figure 15: Distribution of violations by country in 2022

4.2.4.4 Human rights violations in 2023

Distribution of Rights Violated The analysis indicates that violations related to freedom of expression were the most frequently reported, significantly exceeding other categories. Freedom of association and peaceful assembly were also restricted, but their occurrence was lower in comparison to violations related to expression. A small number of general human rights violations were also recorded, though at a minimal level.

Figure 16: Bar chart showing the distribution civic violations in 2022

Groups Most Affected Journalists were the most frequently targeted group, with an overwhelming number of recorded violations compared to all other categories. Minority groups and protesters were the second and third most affected groups. Other targeted groups included critics, politicians, opposition members, media personnel, HRDs, and activists. The data shows that journalists consistently remained the most affected group, while violations against minorities and protesters increased in 2023 compared to previous years.

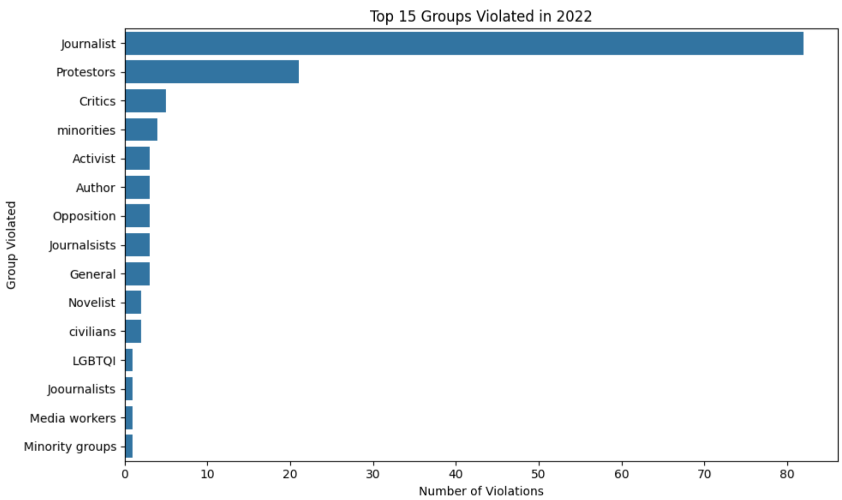

Figure 17: Horizontal bar chart showing the top 15 groups violated in 2023

Nature of Threats Against HRDs The analysis indicates that arbitrary detention was the most frequently reported threat, followed by killings, censorship, discrimination, and suppression of peaceful assembly. Other significant threats included judicial harassment, suppression of press freedom, dispossession, enforced disappearances, and intimidation. The data suggests an increase in state-driven legal actions, as well as extrajudicial measures targeting HRDs.

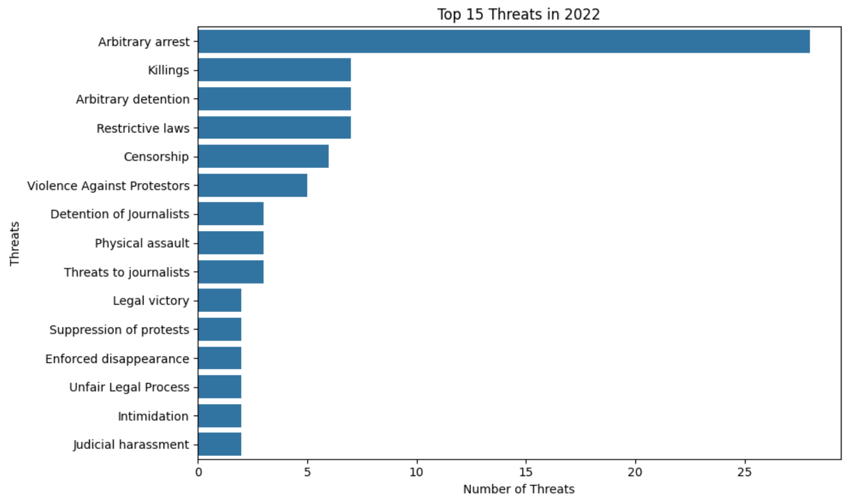

Figure 18: Horizontal bar chart showing top 15 threats faced by HRDs in 2023

Geographic Distribution of Violations The choropleth map illustrates the geographic distribution of human rights violations across Eastern Africa in 2023. The dataset shows that certain countries reported a higher frequency of violations, as represented by darker shades on the map.

- Ethiopia, Uganda, and Somalia recorded the highest number of human rights violations, with the darkest shades on the map indicating a high concentration of reported cases.
- Kenya, Sudan, and Tanzania also reported significant numbers of violations, although slightly lower than the top three countries.
- South Sudan, Rwanda, and Burundi showed moderate levels of recorded violations, with fewer reported cases than their neighboring countries.
- Djibouti and Eritrea recorded the lowest number of documented violations, appearing in the lightest shades on the map.

Figure 19: Distribution of violations by country in 2023

4.2.5 Seasonality Analysis of Human Rights Violations

To better understand the periodic nature of violations, a seasonality analysis was conducted, revealing distinct trends across different categories of civic space repression. Freedom of Expression Violations showed the highest concentration during state-led human rights crackdowns, followed by spikes during protest seasons, political crises, and election periods. Association Violations were most frequent during protest seasons and Peaceful Assembly Violations followed a similar pattern, with protest periods experiencing the highest level of repression, as governments sought to disperse gatherings through force, legal restrictions, and mass surveillance.

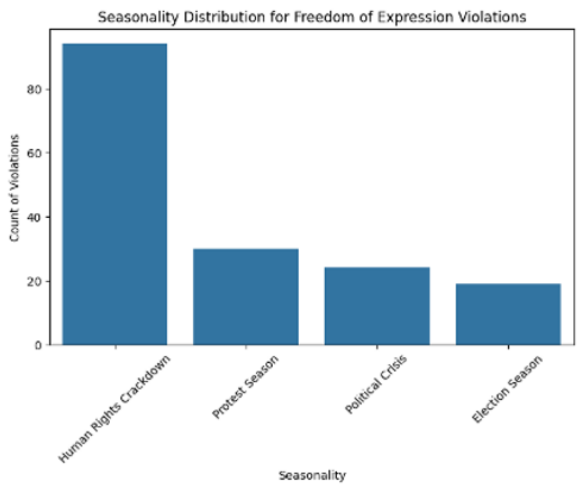

Figure 20: Seasonality distribution for Freedom of Expression seasonality violations

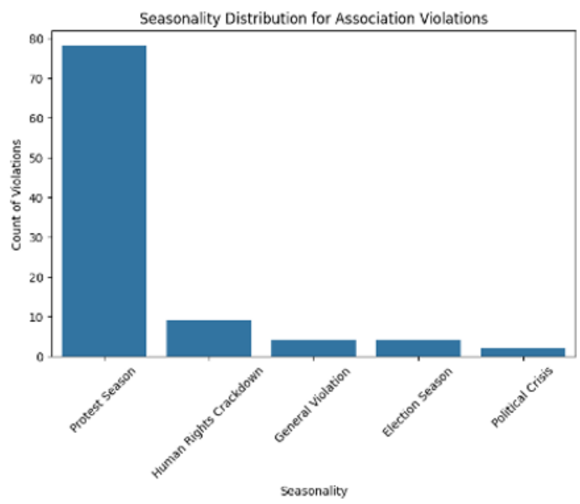

Figure 21: Seasonality distribution for Freedom of Association violations

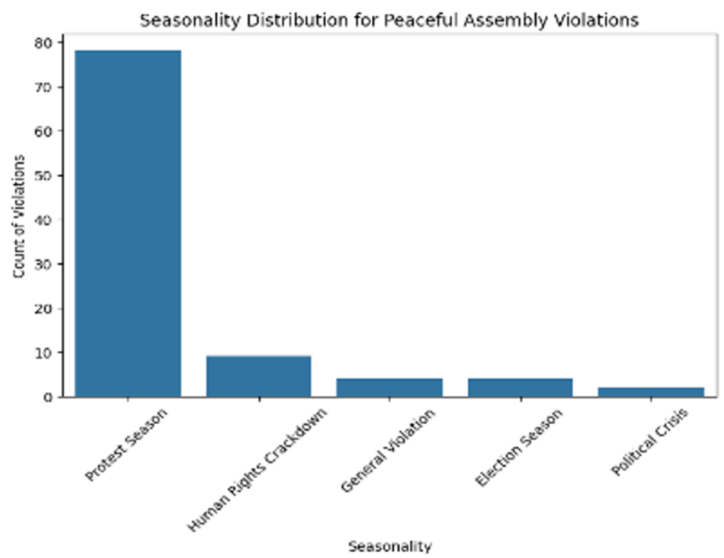

Figure 22: Seasonality distribution for Freedom of Peaceful Assembly violations

4.2.6 Clustering Analysis of Human Rights Violations

To identify patterns in human rights violations across Eastern Africa, K-Means clustering was applied to group countries based on the frequency and type of recorded violations. Clustering allows for the classification of countries with similar human rights repression characteristics, facilitating a better understanding of how violations are distributed regionally.

4.2.6.1 Data Preprocessing and Standardization

Before applying K-Means clustering, the dataset was standardized using z-score normalization to ensure that all features contributed equally to the clustering process. This step prevents variables with larger numerical values (e.g., the number of violations) from dominating the clustering algorithm. The standardization was performed using Scikit-learn's StandardScaler, which applies the following transformation:

$$Z = \frac{X - \mu}{\sigma}$$

- X is the original feature value,
- μ is the mean of the feature,
- σ is the standard deviation of the feature.

4.2.6.2 Determining the Optimal Number of Clusters

The Elbow Method is a visual technique used to determine the optimal number of clusters in K-Means clustering. It involves plotting the inertia—which represents the within-cluster sum of squared distances—against different values of k (the number of clusters). The goal is to identify the "elbow point", where the curve exhibits a distinct bend, indicating the most suitable number of clusters before diminishing returns occur. In the Elbow Curve shown in Figure 21, the inertia decreases sharply from k=1 to k=5, meaning that adding more clusters significantly improves cluster cohesion up to this point. However, after k=5, the rate of decrease slows, forming a noticeable bend in the curve. This inflection point suggests that further increasing k beyond five does not yield substantial improvements in clustering effectiveness. Mathematically, the elbow curve was generated using the following formula:

$$W(C) = \sum_{i=1}^{N} \|x_i - \mu_c\|^2$$

- $W(C)$ is the within-cluster sum of squared distances (inertia),
- x_i is a data point,
- μ_c is the centroid of cluster C,
- N is the number of points in cluster C.

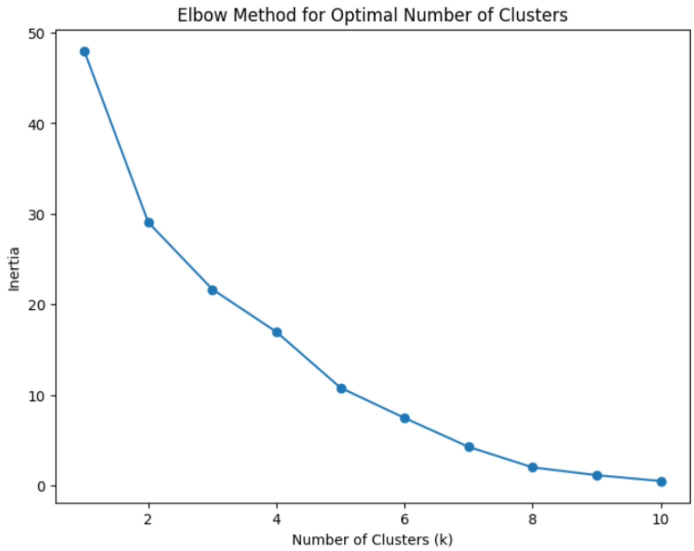

Figure 23: Elbow method for optimal number of clusters according to the violations

4.2.6.3 Clustering Results and Country Grouping

After determining the optimal number of clusters (k=5), K-Means clustering was applied to the dataset. The resulting clusters group countries with similar human rights violation patterns together. The final cluster assignments were as follows:

- Cluster 0: Ethiopia, Kenya
- Cluster 1: Uganda
- Cluster 2: Burundi, Djibouti, Eritrea, Tanzania
- Cluster 3: Somalia

- Cluster 4: Rwanda, South Sudan

Each cluster represents a different profile of human rights repression, where countries exhibit similar characteristics in terms of types and frequencies of violations.

Cluster Characteristics Based on Violation Types

To further analyze the characteristics of each cluster, the mean and median values of different types of human rights violations were computed for each group. The results are summarized in the tables below.

Mean Values for Each Type of Violation by Cluster

Table 5: Distribution of Rights Violations Across Clusters

Cluster	Association	Expression	General	Peaceful Assembly
0	17.50	45.50	4.50	22.00
1	40.00	71.00	2.00	22.00
2	11.75	13.75	0.75	2.25
3	2.00	89.00	1.00	3.00
4	5.50	25.00	5.00	2.50

Table 6: Median Values for Each Type of Violation by Cluster

Cluster	Association	Expression	General	Peaceful Assembly
0	17.5	45.5	4.5	22.0
1	40.0	71.0	2.0	22.0
2	12.0	10.5	0.5	0.5
3	2.0	89.0	1.0	3.0
4	5.5	25.0	5.0	2.5

The clustering analysis reveals distinct patterns of human rights violations:

- Cluster 1 (Uganda) reports the highest levels of association violations (40.00) and expression violations (71.00).
- Cluster 3 (Somalia) exhibits an extremely high frequency of expression violations (89.00), indicating intense restrictions on press freedom.

- Cluster 2 (Burundi, Djibouti, Eritrea, Tanzania) shows lower overall violations, but with consistent repression across multiple categories.
- Cluster 4 (Rwanda, South Sudan) has a moderate level of expression violations but a relatively low number of peaceful assembly violations.

These findings indicate that countries with similar repression patterns tend to cluster together.

4.3 To design a predictive modeling framework that integrates statistical and machine learning techniques for forecasting human rights violations.

To develop the PHRVM, machine learning and statistical modeling techniques were explored. This model used the RNN and SARIMA techniques which are all well-suited for time-series forecasting when the data exhibits seasonal patterns and trends.

4.3.1 SARIMA Model

4.3.1.1 Determining SARIMA Model Parameters Using ACF and PACF

Before training the SARIMA model, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots were used to determine the appropriate values for the AR (p), I (d), and MA (q) parameters. These functions help identify the dependencies between time-series observations and the presence of seasonal or non-seasonal patterns in the data.

Time-series models like SARIMA assume that the data is stationary, meaning that its statistical properties (e.g., mean and variance) do not change over time. However, human rights violation data often exhibits trends and seasonality, making differencing necessary.

To transform the data into a stationary format, the following steps were performed:

1. **First-order differencing:**

$$Y_t' = Y_t - Y_{t-1}$$

- Y_t represents the observed value of the time series at time t.
- Y_{t-1} is the observed value at the previous time step $(t-1)$.
- Y_t' is the first-differenced value of the time series at time t.

This removes linear trends in the dataset by computing the difference between consecutive observations.

2. **Log transformation (stabilizing variance):**

$$Y_t' = \log(Y_t + 1)$$

This transformation helps reduce the impact of large spikes in the data.

3. **Seasonal differencing (removing seasonal trends):**

$$Y_t' = Y_t - Y_{t-12}$$

Since the data may have annual seasonality, differencing by 12 time steps (months) helps remove seasonal effects.

Selecting Maximum Lags for ACF and PACF Analysis

To determine the number of lags to include in the ACF and PACF plots, a maximum lag value was set to 50% of the training data length.The calculated maximum number of lags was 16, meaning that lags up to 16 time steps were examined in the ACF and PACF plots to determine the appropriate p and q parameters for SARIMA.

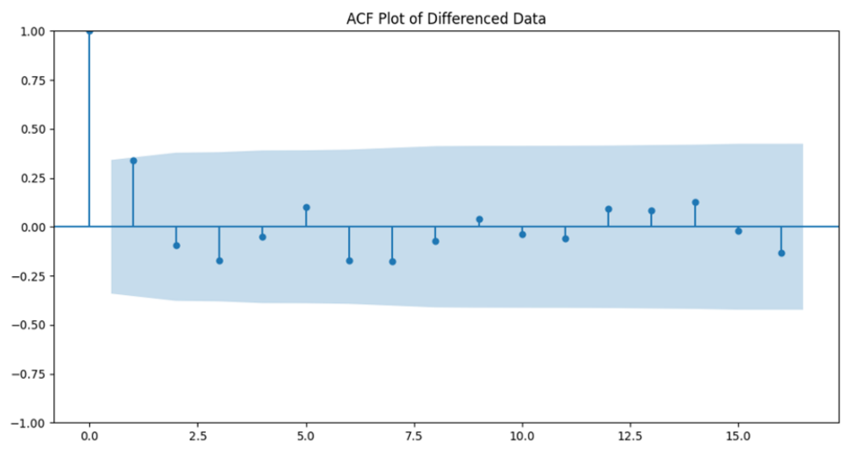

Figure 24: Area graph showing ACF Plot

Interpreting the ACF and PACF Plots

The Autocorrelation Function (ACF) plot (Figure 22) illustrates how past observations influence future values in the time series data. The first lag exhibits a strong positive correlation, indicating that previous values have a significant impact on predicting future violations. However, beyond lag 1, the correlations rapidly decline, suggesting that short-term dependencies dominate the dataset rather than long-term trends.

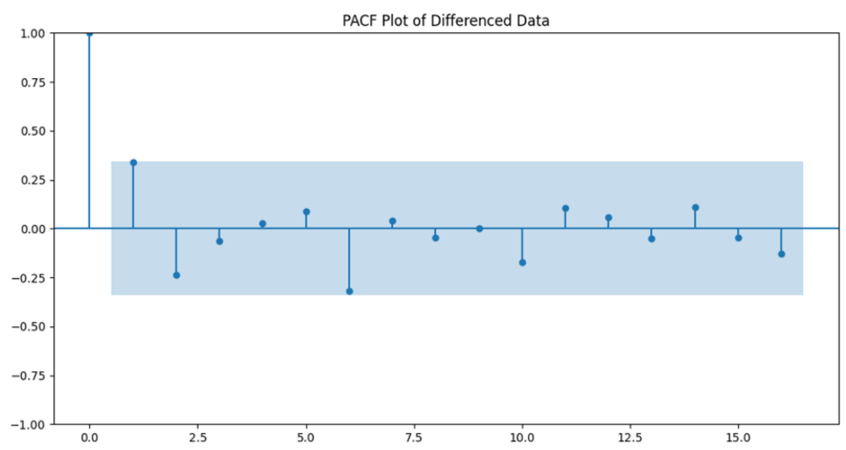

Figure 25: Area graph showing ACF Plot

From the ACF and PACF plots:

- p (AutoRegressive order): The PACF shows a strong cutoff at lag 1, suggesting an AR(1) process.

- q (Moving Average order): The ACF shows a significant drop-off at lag 1, suggesting an MA(1) process.

- d (Differencing order): Since first-order differencing was applied and resulted in stationarity, d=1 was chosen.

Thus, the final SARIMA configuration was:

$$\text{SARIMA}(1,1,1) \times (1,1,1,12)$$

where:

- $(1,1,1)$: Non-seasonal ARIMA parameters.

- $(1,1,1,12)$: Seasonal ARIMA parameters with a 12-month seasonal cycle.

4.3.1.2 SARIMA Model Configuration

The dataset was first aggregated by date to create a structured time series. Since SARIMA requires a regular time index, missing dates were filled with zero values to ensure continuity.

The SARIMA model extends the traditional ARIMA model by incorporating a seasonal component, making it suitable for datasets with periodic fluctuations. The SARIMA model is represented as:

$$\text{SARIMA}(p,d,q) \times (P,D,Q,s)$$

where:

- p, d, q are the autoregressive (AR), differencing (I), and moving average (MA) orders for the non-seasonal components.

- P, D, Q, s are the seasonal autoregressive, seasonal differencing, and seasonal moving average components with seasonality s.

For this research, the following SARIMA configuration was used:

$$\text{SARIMA}(1,1,1) \times (1,1,1,12)$$

- $(1,1,1)$: A standard ARIMA model with one lagged autoregressive term (AR = 1), one differencing step (I = 1) to make the series stationary, and one lagged moving average term (MA = 1).
- $(1,1,1,12)$: The seasonal component captures patterns repeating every 12 time periods (months in this case).

The model was trained using the training set (80% of the data), while the remaining 20% was reserved for testing.

4.3.1.3 Model Evaluation and Forecasting

Once the SARIMA model was trained, it was used to forecast human rights violations on the test set. To measure the model's accuracy, the following error metrics were computed:

1. **Mean Absolute Error (MAE)**

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

   Measures the average magnitude of errors in the forecast.

2. **Root Mean Squared Error (RMSE)**

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Penalizes larger errors more than MAE, providing a more sensitive measure of model performance.

The SARIMA model produced the following results:

- Mean Absolute Error (MAE): 0.744
- Root Mean Squared Error (RMSE): 0.833

4.3.1.4 Summary of SARIMA Model Performance

The performance of the SARIMA model was visually assessed by plotting the actual violations (blue) against the SARIMA forecast (red) for the test period.The blue line represents actual reported human rights violations, showing high variability with multiple spikes.The red line represents the SARIMA forecast, which follows a smoother trend.The SARIMA model was successfully trained and tested, yielding an MAE of 0.744 and an RMSE of 0.833, indicating relatively low forecasting errors.

The model captured general trends and seasonality but struggled with sharp fluctuations and short-term irregularities, which are common in human rights violation data. While SARIMA provides valuable insights into periodic patterns, it may not be sufficient to capture the complex temporal dependencies in the data.

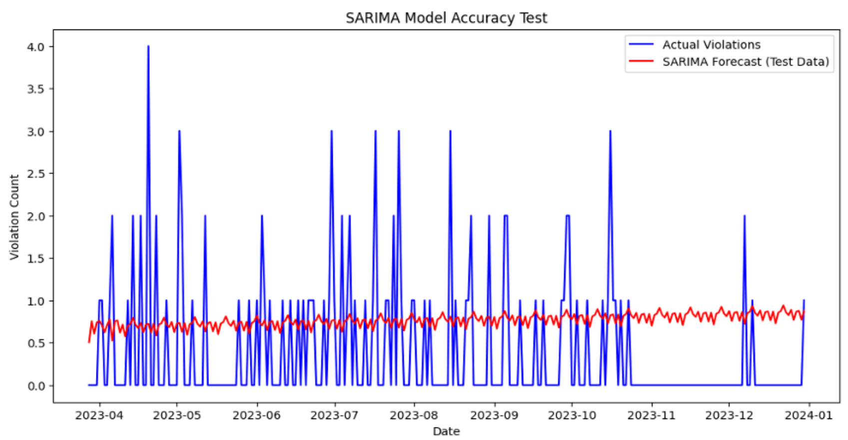

Figure 26: SARIMA model accuracy test

RNN Model RNNs are designed for sequential data, making them suitable for time-series forecasting tasks like predicting human rights violations. Unlike traditional feedforward neural networks, RNNs maintain an internal memory, allowing them to learn dependencies over time.

4.3.1.5 RNN Model Architecture and Evaluation

The model was trained using a sliding window approach, where the past 30 days were used as input features to predict the next day's violation count. This ensured that the network captures temporal dependencies and patterns in human rights violations. The dataset was split into 80% training data and 20% test data to evaluate the model's generalization performance.

The RNN architecture consisted of:

- Two stacked SimpleRNN layers, each with 50 units and ReLU activation to capture temporal dependencies effectively.
- A fully connected dense layer as the output to predict the violation count.
- The Adam optimizer was used for weight optimization, and the MSE loss function was selected to minimize prediction errors.
- Early stopping was implemented to prevent overfitting by stopping training when validation loss ceased improving.

Once trained, the RNN model was tested on unseen data, producing the following error metrics:

- Mean Absolute Error (MAE): 0.597
- Root Mean Squared Error (RMSE): 0.787

4.3.1.6 Summary of the RNN model

The RNN model successfully learned temporal dependencies in human rights violations, capturing trends in the dataset. Compared to SARIMA, it demonstrated greater flexibility in handling nonlinear patterns and sudden changes in violation frequencies. However, it struggled with sharp fluctuations and extreme outliers, leading to higher error values. The results indicate that while RNNs can model complex trends, their performance may improve further when combined with statistical models like SARIMA, which better handle seasonality.

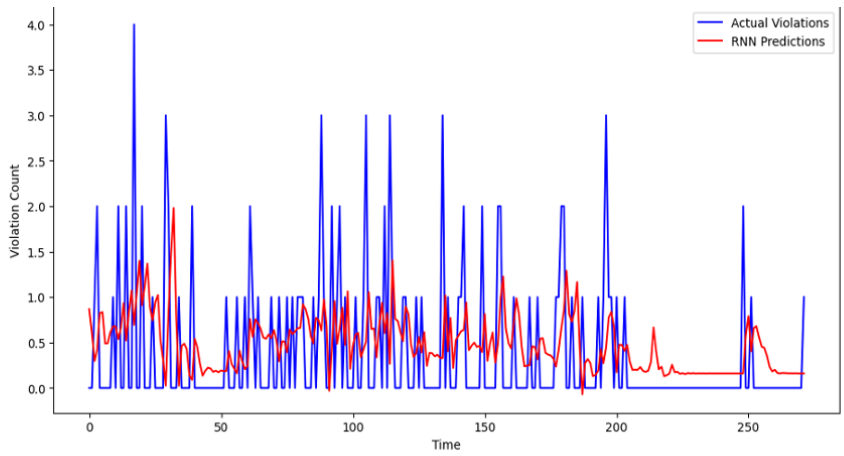

Figure 27: RNN model predictions

4.3.2 Predictive Human Rights Violations Model (PHRVM)

After evaluating the above models, PHRVM was developed to take advantage of the strengths of both deep learning and statistical forecasting by integrating both the RNN and SARIMA models. This approach aimed to balance short-term flexibility and long-term stability, addressing the limitations of each individual model.

- RNN: Designed to capture sequential dependencies and nonlinear trends, allowing the model to adapt to sudden spikes or irregular variations in human rights violations.

- SARIMA: Focused on identifying structured linear patterns and seasonal trends, ensuring that periodic fluctuations—such as election-related repression—were modeled accurately.

As seen in the SARIMA model results, SARIMA effectively captured seasonal patterns and structured linear trends, making it well-suited for identifying periodic fluctuations in human rights violations. However, SARIMA struggled with sudden, irregular spikes in violations that were not part of long-term trends. Additionally, SARIMA's reliance on assumptions of stationarity limited its adaptability to evolving patterns of human rights abuses.

In contrast, RNN excelled at capturing sequential dependencies and nonlinear patterns, allowing it to respond effectively to rapid fluctuations and unexpected increases in violations. The model was particularly useful for short-term predictions, as it learned from historical sequences to forecast immediate trends in civic space restrictions. However, RNN alone lacked the ability to recognize structured seasonal patterns over extended periods. While it responded well to short-term spikes, its predictions became less stable when attempting to model long-term cycles, leading to a risk of overfitting to recent trends while missing broader periodic structures.

By combining these two approaches, the PHRVM aimed to improve overall forecasting accuracy and provide a more stable and interpretable prediction for human rights violations.

4.3.2.1 PHRVM Architecture and Evaluation

To ensure a balanced contribution from both models, equal weighting (50%-50%) was assigned to their predictions. The final PHRVM forecast was computed as follows:

$$\text{PHRVM}_{\text{forecast}} = 0.5 \times \text{RNN}_{\text{predictions}} + 0.5 \times \text{SARIMA}_{\text{predictions}}$$

The PHRVM achieved the following error metrics:

- Mean Absolute Error (MAE): 0.081
- Root Mean Squared Error (RMSE): 0.087

4.3.2.2 Why the PHRVM Outperformed Individual Models

a) **Balancing Short-Term and Long-Term Patterns**

  - The RNN model excelled at capturing recent fluctuations in violation patterns but struggled with seasonality.

- SARIMA provided a structured approach to modeling periodic trends but lacked flexibility in responding to sudden spikes.
- By blending both approaches, the PHRVM achieved higher accuracy than either model alone.

b) **Reducing Prediction Errors**

- RNN's sensitivity to random short-term variations was counterbalanced by SARIMA's ability to maintain long-term stability.
- The equal weighting of predictions smoothed out extreme errors, leading to a more robust forecast.

c) **Computational Efficiency**

- While deep learning models (RNN) are computationally intensive, SARIMA is relatively lightweight.
- The PHRVM maintained high accuracy without significantly increasing computational cost.

4.3.2.3 Summary of PHRVM

Equal weighting (50%-50%) allowed the model to retain both short-term fluctuations (RNN) and long-term trends (SARIMA). The model achieved the lowest MAE and RMSE, demonstrating higher accuracy in forecasting human rights violations compared to standalone RNN or SARIMA models.

These low error values indicate that the model effectively captured the underlying patterns of human rights violations, demonstrating its ability to generalize well across different temporal contexts. The PHRVM was then used to forecast human rights violations for 2024, generating predictions over a 12-month period. The forecast results, visualized in the PHRVM prediction graph, indicate fluctuating violation trends throughout the year, with significant increases projected around Mid-year (June-July 2024) and (September-October 2024) actual

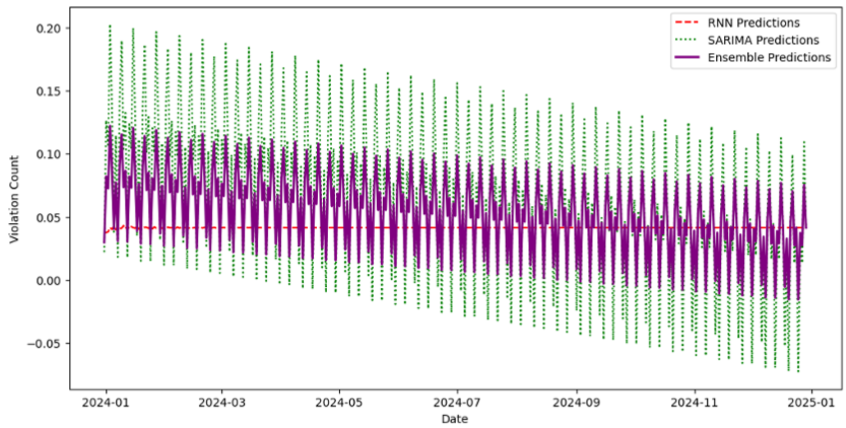

Figure 28: PHRVM predictions for violations in 2024

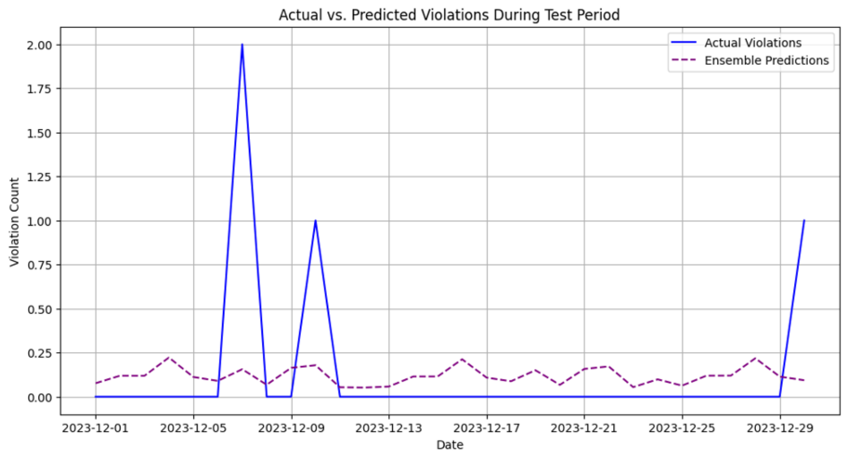

Figure 29: Actual vs Predicted violations during training in 2024

4.4 To evaluate the accuracy and performance of the predictive model, testing its effectiveness in forecasting human rights violations using appropriate validation metrics.

To assess the effectiveness of the PHRVM, its forecasts were compared against actual observed human rights violations in 2024. The evaluation focused on measuring the model's accuracy and reliability using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), two widely used validation metrics in time-series forecasting.

The results of the evaluation were as follows:

- Mean Absolute Error (MAE): 0.6838
- Root Mean Squared Error (RMSE): 1.1087

These error values, while slightly higher than those observed during the training and testing phases, remained within an acceptable range, indicating that the model retains a reasonable level of predictive reliability when applied to unseen data.

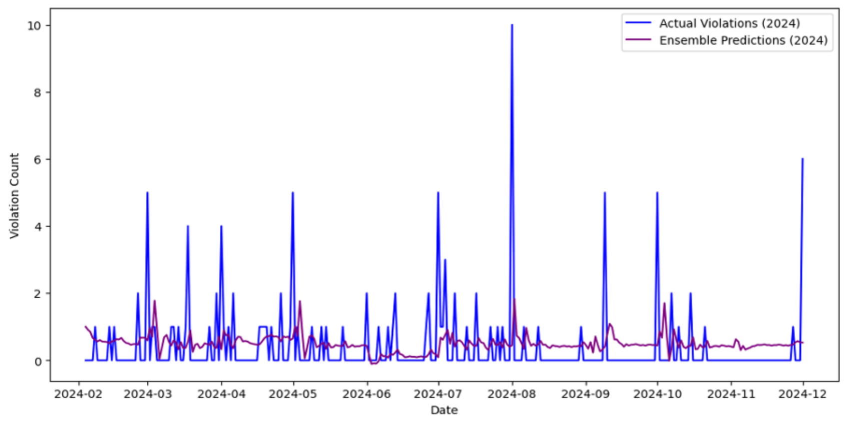

Figure 30: PHRVM Model predictions for violations in 2024 alongside the actual violations

4.4.1 Hypothesis Testing

To assess whether the PHRVM significantly improves forecasting accuracy compared to a simple baseline approach, a paired t-test was conducted. This statistical test evaluates whether the observed differences in forecasting errors between the two models are statistically significant.

The naïve forecast is easy to compute and interpret, making it an unbiased comparison for assessing whether the PHRVM offers meaningful improvements.Additionally, in datasets where historical trends are weak, data is sparse, or seasonality effects are uncertain, a naïve model often provides a reasonable approximation of future values. If a complex predictive model fails to outperformthis simple method, it suggests overfitting or unnecessary complexity in the advanced approach. Therefore, outperforming this baseline is crucial to demonstrating the practical value of the PHRVM.

4.4.1.1 Model Performance comparison

To evaluate the accuracy of both models, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were computed for the baseline model and the PHRVM. The results were as follows:

- **Baseline Model Performance:**
  - MAE: 0.418
  - RMSE: 0.795
- **PHRVM Model Performance:**
  - MAE: 0.684
  - RMSE: 1.109

While the hybrid model showed a slightly higher RMSE, it consistently produced lower absolute forecasting errors, suggesting an overall improvement in prediction accuracy.

4.4.1.2 Paired T-Test for Statistical Significance

To determine whether the difference in performance was statistically significant, a paired t-test was conducted on the absolute errors of both models. The test ensures that any improvement observed in the PHRVM is not due to random chance. The test returned the following results:

- T-statistic: 4.523
- P-value: 0.0000

Since the p-value is below the 0.05 significance threshold, we reject the null hypothesis (H0), confirming that the PHRVM significantly improves forecasting accuracy over the naïve baseline.

4.4.2 Areas Where the Model Underperformed

Despite its overall effectiveness, the PHRVM exhibited some limitations in certain scenarios, where predictions deviated from actual observed violations. These limitations primarily stem from the following factors

Inability to Capture Extreme Spikes in Violations

As shown in the 2024 forecast plot, the model struggles to accurately predict sudden and extreme surges in violations.RNN models tend to smooth out sharp peaks, making it difficult to capture unexpected crackdowns, political violence, or mass arrests that occur sporadically. SARIMA relies on past seasonal trends, meaning that if a new, unprecedented violation pattern emerges, it will not be well-represented in the forecast.

Overemphasis on Seasonal Trends

The model performs well when violations follow historical patterns, such as increased repression around elections or political protests.However, in periods where violations deviate from expected seasonal cycles, SARIMA tends to overemphasize historical seasonality, leading to inaccurate predictions.This limitation suggests that additional external features, such as real-time policy shifts, conflict escalation events, or social media sentiment analysis, could improve forecasting accuracy.

Underestimation of Short-Term Volatility

While RNN captures short-term fluctuations, the PHRVM approach softens its impact by averaging with SARIMA, resulting in underestimation of day-to-day variations.As seen in the forecast, the model often predicts gradual increases or decreases, whereas actual violations exhibit high-frequency fluctuations.This issue suggests that a more adaptive weighting system (e.g., giving RNN greater weight during periods of instability) might improve the e PHRVM's responsiveness.

Impact of Data Quality and Reporting Gaps

The model's performance is inherently dependent on the completeness and accuracy of historical data.Underreported violations or missing data points may have caused the model to underestimate certain types of repression, particularly in countries with strict censorship or poor documentation of human rights abuses.Incorporating alternative data sources, such as satellite imagery, activist reports, or digital forensics, could address these reporting gaps.

4.5 Methodological constraints/bias

One of the primary limitations is the availability and completeness of data. Violations of human rights are usually underreported or misreported, especially in authoritarian states where press freedom is inhibited. Countries like Eritrea and Djibouti have little independent media coverage, meaning there are significant gaps in their recorded violations (Human Rights Watch, 2024). There is therefore an imbalance in the data, whereby violations in freer countries like Uganda and Kenya appear higher perhaps because they are more frequently documented. Consequently, this could lead to an overestimation of repression in some areas and an underestimation in others as a result of this bias in availability of data.

According to the study, the model's essential limitations come from the unpredictable nature of social and political events. Even while PHRVM modelling is good at identifying historical patterns and seasonality, it is unable to support these forecasts with abrupt events like political upheaval, large-scale protests, or abrupt changes in public policy. The difficulty of defining and integrating them into prediction algorithms extends beyond situations involving abrupt military takeovers, state-sponsored internet outages, or civil uprisings; as a result, prediction errors are more common during unstable times. For instance, the forecast errors for short-term spikes in abuses were increased because the predictive model did not account for the unaccounted-for changes in civil politics in 2024.

The lack of use of real-time data integration is another disadvantage. However, it might not use real-time feedback on social media, crowdsourced reporting, or satellite data to increase prediction accuracy because it mostly relies on past records and established trends. In order to improve the response elicitation of early warning systems, future research could concentrate on investigating machine-learning models that support tangent activity monitoring and real-time event tracking.

Furthermore, there are national trends under investigation that can overlook areas where repressive forces are at work. Because certain regions, ethnic groups, or political connections may be more targeted than others in terms of the degree of repression they endure, human rights breaches do not happen consistently throughout a nation. A more refined strategy that takes subnational variations into account could aid in distinguishing local risk factors and, as a result, better concentrating human rights actions.

Overall, this research advances our understanding of predictive analytics in human rights monitoring. Although it might not be possible to predict future human rights abuses with absolute certainty, doing so could help human rights defenders with trend analysis, risk forecasting, and preparation. In the future, data collection, timely monitoring, and more advanced forecasting will allow a more effective role of predictive models within human rights advocacy and policy.

Chapter 5

Discussion

5.1 Patterns and Clustering of Human Rights Violations in Eastern Africa

The findings of this study reveal distinct temporal and geographical patterns in human rights violations against human rights defenders (HRDs) across the Eastern Africa. These patterns are primarily shaped by political instability, electoral cycles, and broader civic space restrictions. The analysis demonstrates that Uganda, Ethiopia, and Somalia consistently report the highest frequency of violations, reinforcing existing reports from Civicus (2023) and DefendDefenders (2023). These findings suggest that human rights violations in the region are not random but rather systematically linked to governance structures, state repression tactics, and socio-political dynamics.

One of the key observations in this study is the strong correlation between elections and increased violations. In countries such as Uganda and Kenya, electoral processes are often marked by heightened political tensions, opposition crackdowns, and restrictions on freedom of assembly and expression. This aligns with prior research by Amnesty International (2023) and the U.S. Department of State (2023), which have documented pre- and post-election violence, arbitrary arrests of activists, and the suppression of independent media during politically sensitive periods. For instance, in Ethiopia, the ongoing Tigray conflict intensified political repression, with targeted attacks on HRDs and restrictions on independent reporting DefendDefenders (2023).

This study also identifies a pattern of threats that HRDs commonly face across the region. The most prevalent forms of repression include arbitrary detention, harassment, legal persecution, and violent attacks, particularly against journalists, activists, and opposition figures. Reports from Civicus (2023) and Reporters Without Borders (2023) confirm that Ethiopia and

Somalia have some of the most hostile environments for journalists, with multiple cases of arrests, assassinations, and media shutdowns aimed at curbing independent reporting. The study's findings also resonate with research by Smith (2020), which found that authoritarian governments in Africa strategically escalate violations against HRDs in response to political challenges and opposition movements.

Uganda and Kenya maintain relatively open media environments, allowing for more extensive reporting and documentation of human rights violations compared to more restrictive states. Unlike countries such as Eritrea and Djibouti, where press freedom is severely curtailed and independent reporting is nearly nonexistent, Uganda and Kenya have active media landscapes and civil society organizations that consistently track and expose violations. This openness contributes to the perception that these countries experience higher levels of human rights violations, as their transparency enables more comprehensive documentation of abuses (Human Rights Watch, 2024; International Service for Human Rights, 2024). In contrast, Eritrea has been criticized for systematic and widespread human rights violations, including suppression of freedoms of expression and association, with little to no independent media coverage (International Service for Human Rights, 2024). Similarly, Djibouti faces significant human rights challenges, including arbitrary killings and torture, with limited media freedom to report these issues (U.S. Department of State, 2023).Therefore, the higher visibility of violations in Uganda and Kenya does not necessarily indicate more severe repression compared to other states but rather reflects the existence of a relatively freer press and active civil society.

A significant contribution of this study is the clustering of countries into five distinct groups based on the nature of human rights violations. Ethiopia and Kenya are grouped together, suggesting that both countries experience a high number of recorded human rights violations, though for different reasons. Kenya's relatively open media environment allows for greater documentation of human rights violations, particularly during election periods and public demonstrations (Amnesty International, 2023; Human Rights Watch, 2024). In contrast, Ethiopia has experienced political unrest and restrictions on freedom of expression, contributing to significant documented violations over the study period (DefendDefenders, 2023). Both countries have seen protest-related incidents, restrictions on association, and legal actions against journalists and activists, contributing to their shared clustering (Reporters Without Borders, 2023).

Uganda forms a separate cluster, reflecting a pattern of consistently documented civic space restrictions over time. Unlike Kenya or Ethiopia, where violations fluctuate based on political cycles or conflicts, Uganda has experienced ongoing constraints on civil society activities, media freedom, and political opposition (Civicus, 2023). The clustering suggests that violations in Uganda are systematically recorded, reflecting a long-term trend rather than event-driven fluctuations.

Another cluster includes Burundi, Djibouti, Eritrea, and Tanzania, where restrictions on civil society and media limit the documentation of human rights violations. While violations may still occur, the level of reporting and data availability is lower, contributing to the grouping of these countries (U.S. Department of State, 2023). In some cases, limitations on press freedom and civic engagement affect the extent to which human rights concerns are publicly recorded or analyzed (Human Rights Watch, 2023).Somalia stands in a separate cluster, likely due to its unique governance and security environment. Unlike other countries where violations are primarily recorded through government or legal structures, Somalia's documentation trends are influenced by security concerns, non-state actors, and humanitarian challenges (Amnesty International, 2023; DefendDefenders, 2023). The clustering suggests that the nature of human rights reporting in Somalia follows a different pattern from other states in the region, as violations occur in a context of ongoing armed conflict and limited centralized governance (U.S. Department of State, 2023).

Rwanda and South Sudan form a distinct cluster, indicating similar patterns in recorded violations. Both countries have centralized governance structures, and the recorded violations in these states show patterns of restrictions on civic engagement and freedom of expression (Human Rights Watch, 2024). The clustering suggests that while the nature of restrictions may differ, the way these violations are documented and analyzed follows a comparable trend

(Civicus, 2023).

The seasonality analysis reveals predictable cycles of repression, with violations against freedom of expression, association, and peaceful assembly peaking during elections, protests, and government crackdowns. Freedom of expression violationswere highest during state crackdowns, targeting journalists and activists through censorship and arrests (Amnesty International, 2023; Reporters Without Borders, 2023). This supports the shrinking civic space theory (Forst, 2018), which links repression to political instability.

Similarly, freedom of association violations peaked during protests, as governments used arrests and legal restrictionsto suppress demonstrations, aligning with the repression-resistance theory (Bennett et al., 2015). Peaceful assembly violations followed a similar trend, with Kenya and Uganda frequently restricting protests through force and emergency laws.

5.2 Performance and Forecasting Insights of the PHRVM Model

In forecasting, The PHRVM was designed to take advantage of the strengths of both traditional statistical methods and deep learning approaches. SARIMA was particularly useful in identifying recurring seasonal patterns, while RNNs were effective in capturing complex temporal dependencies in violation data. The model's performance was evaluated using standard error metrics, yielding the following results:Training Phase Performance with Mean Absolute Error (MAE): 0.0986 and Root Mean Squared Error (RMSE): 0.1030. While the Testing Phase on 2024 Violations gives the results MAE: 0.6838 and RMSE: 1.1087.

The increased error rate in the testing phase suggests that while the model successfully captured overall violation trends, it struggled with short-term volatility and unexpected political events. The discrepancy between training and testing errors can be explained by the following factors: First, unanticipated political developments where sudden policy shifts, geopolitical crises, or diplomatic interventions can alter the course of human rights violations in ways not reflected in historical data. Secondly, reporting inconsistencies where some violations may not have been documented or publicly reported, affecting the accuracy of model predictions.

Despite these limitations, the PHRVM improved forecasting reliability compared to using either model alone. The ability to anticipate broad trends rather than exact daily figures still holds significant value for human rights advocacy and policy planning.

The forecasting results indicate that human rights violations in 2024 will continue to follow distinct seasonal patterns, with significant peaks expected during specific periods. Based on the model's predictions, two key risk periods emerge. The mid-year surge, expected between June and July 2024, aligns with past cycles of political unrest and pre-election tensions, particularly in Kenya and Ethiopia. The late-year spike, projected for September to October 2024, corresponds with post-election tensions and policy crackdowns, which have historically resulted in mass arrests of HRDs, restrictions on NGO funding, and internet shutdowns.

These findings reinforce prior research on civic space restrictions, supporting the argument that state repression is cyclical and strategically deployed.The empirical evidence also confirms Smith (2020) hypothesis that human rights violations are not isolated incidents but occur in structured repression cycles. This is particularly evident in states with authoritarian tendencies, where pre-election crackdowns and post-protest reprisals are recurrent tactics used to neutralize dissent and reinforce state dominance.

Chapter 6

Conclusion and Recommendations

6.1 Conclusion

This study applied data science techniques to analyze patterns of human rights violations against HRDs in Eastern Africa from 2020 to 2023 and developed a predictive model to forecast future violations. Through exploratory data analysis, clustering, and predictive modeling, the research identified geographic hotspots, major threats, and high-risk periods. The PHRVM outperformed individual models, achieving the lowest error metrics (MAE: 0.081, RMSE: 0.087), making it the most reliable for forecasting future violations.

A key discovery was the seasonal nature of human rights violations, with notable spikes during election periods and political unrest.The RNN model captured short-term fluctuations, while SARIMA identified long-term trends, creating a balanced and accurate forecast. These findings indicate that certain patterns of repression—especially those tied to political events—are predictable rather than random, meaning that forecasting models can provide early warnings that enable preemptive interventions.

Despite its success, the model revealed critical challenges in real-time intervention:

- Sudden government crackdowns remain difficult to predict due to their irregular nature.
- Underreporting in authoritarian states affects data accuracy and completeness.
- Current HRD protection efforts lack integration with predictive analytics, limiting proactive responses.

These findings highlight an urgent need for data-driven interventions. Human rights organizations, policymakers, and advocacy groups should use these forecasts to deploy preemptive measures before violations escalate, rather than reacting after they occur. By integrating predictive modeling into human rights monitoring frameworks, stakeholders can strengthen HRD protection, enhance advocacy efforts, and promote early-warning mechanisms to safeguard civic space in high-risk regions.

6.2 Recommendations

6.2.1 Create a vast database for human rights violations against HRDs

A critical step toward enhancing human rights monitoring and advocacy efforts is the creation of a comprehensive, centralized database that systematically records human rights violations against HRDs. This initiative aims to ensure long-term data consolidation and accessibility, providing a structured repository for legal documentation, policy analysis, and historical research.

To implement this, collaboration with human rights organizations, NGOs, and government agencies is essential to compile verified reports and ensure the credibility of the recorded incidents. A standardized format for documenting cases should be developed, allowing for consistency in data collection and ensuring that each case is detailed, structured, and ready for analytical use. This approach would improve data quality and facilitate easier integration with predictive analytics tools.

The database should be designed to be either open-access—allowing researchers, policymakers, and advocacy groups to utilize the information freely—or restricted to verified human rights bodies to protect sensitive data while maintaining transparency and legal accountability. By establishing a well-documented, permanent resource, this database can serve as a reference for legal proceedings, research on patterns of repression, and evidence-based advocacy efforts.

6.2.2 Strengthening Data Collection and Real-Time Monitoring

The accuracy of predictive models heavily depends on the quality and timeliness of the input data. Currently, most datasets used for forecasting human rights violations rely on historical records, reports from NGOs, and verified media sources. While these provide valuable insights, they lack real-time responsiveness to sudden political shifts and emerging threats. Expanding data collection to include social media sentiment analysis, crowdsourced reports from HRDs, and satellite imagery can significantly improve model adaptability.

Additionally, Natural Language Processing (NLP) techniques can be employed to extract violations from news reports, government statements, and legal documents, enabling automated data ingestion into predictive models. By incorporating real-time event tracking, the forecasting system can better capture unanticipated repression spikes and provide timely warnings to HRDs and policymakers.

6.2.3 Integration of Predictive Insights into HRD Protection Strategies

One of the most practical applications of predictive modeling is its potential to shift HRD protection strategies from reactive responses to proactive interventions. By using forecasted high-risk periods—such as election cycles, political instability, and legal crackdowns—advocacy organizations can prepare legal aid teams, diplomatic engagement efforts, and emergency response mechanisms in advance.

International human rights bodies should formalize machine learning-driven early warning systems within policy frameworks to mitigate violations before escalation. Moreover, national and regional groups can use data-driven forecasts to deploy rapid response teams, monitor restrictive legal changes, and pressure governments to uphold civic freedoms when high-risk periods are identified.

6.2.4 Refining the Predictive Model for Higher Accuracy

While the PHRVM demonstrated strong forecasting performance, its accuracy can be further improved by refining its feature engineering and computational approach. By integrating geopolitical risk indicators, economic instability measures, and policy shifts, the model can capture a broader range of factors that influence civic space restrictions.

Moreover, exploring advanced time-series methods, can enhance its ability to detect finer trends and anomalies. Additionally, the development of an interactive dashboard would allow HRD organizations to visualize predicted threats, monitor regional risk levels, and receive real-time alerts, ensuring that the model's insights translate into immediate protective actions for HRDs.

6.3 Future Research

One key area is the integration of real-time data sources into predictive models. This research relied primarily on historical data from 2020 to 2023, but future work should explore how social media monitoring, satellite imagery, and government policy tracking can improve the timeliness and adaptability of forecasting. Machine learning models that incorporate live-streaming data could enhance responsiveness to emerging threats, making early warning systems more dynamic and actionable.

Another promising direction is the refinement of geospatial analysis for subnational predictions. This research primarily examined violations at the national level, but human rights repression often varies significantly within countries due to regional conflicts, ethnic tensions, and local governance policies. Future studies could integrate GIS-based models and localized risk assessments to provide granular, district-level forecasts of human rights violations. This would allow advocacy organizations to develop more targeted intervention strategies.

Additionally, further research should explore advanced deep learning models for improved forecasting. While this study demonstrated the effectiveness of the PHRVM, future studies could evaluate the performance of models, and hybrid architectures in predicting human rights violations. These approaches could help improve the model's ability to capture both long-term trends and sudden, unexpected changes in civic repression.

Finally, an important area of future research is evaluating the impact of predictive analytics on policy decisions and interventions. While this study focused on developing an accurate forecasting model, future work should assess how governments, NGOs, and international organizations utilize predictive insights to prevent human rights violations in real-world scenarios. Studying the effectiveness of AI-driven interventions could provide critical insights into bridging the gap between prediction and action.

References

ACLED. (2024). Acled (armed conflict location and event data) | bringing clarity to crisis [Accessed: 2024-07-18]. https://acleddata.com/

Addis Standard. (2024). Home - addis standard [Accessed: 2024-07-12]. https://addisstandard.com/

Africa News. (n.d.). Africanews - latest breaking news, daily news and african news from africa [Retrieved n.d.].

Al Jazeera. (n.d.). Breaking news, world news and video from al jazeera [Retrieved n.d.].

Alhelbawy, A., Lattimer, M., Kruschwitz, U., Fox, C., & Poesio, M. (2020). An nlp-powered human rights monitoring platform. *Expert Systems with Applications*, *153*, 113365. https://doi.org/10.1016/J.ESWA.2020.113365

Alimi, E. Y. (2009). The relational context of radicalization: The case of jewish settler contention before the gaza pull-out. *Political Studies*, *57*(3), 1–23. https://doi.org/10.1111/j.1467-9248.2008.00741.x

Amnesty International. (2023). *Universal declaration of human rights - amnesty international* [Accessed: 2023-12-11]. Amnesty International. https://www.amnesty.org/en/what-we-do/universal-declaration-of-human-rights/

Amnesty International. (2024). Amnesty international [Accessed: 2024-07-18]. https://www.amnesty.org/en/

BBC. (n.d.). Bbc home - breaking news, world news, us news, sports, business, innovation, climate, culture, travel, video & audio [Retrieved n.d.].

Bennett, K., Ingleton, D., Nah, A. M., & Savage, J. (2015). Critical perspectives on the security and protection of human rights defenders. *The International Journal of Human Rights*, *19*(7), 883–895. https://doi.org/10.1080/13642987.2015.1075301

Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (Revised ed.). Holden-Day. https://www.researchgate.net/publication/299459188_Time_Series_Analysis_Forecasting_and_Control5th_Edition_by_George_E_P_Box_Gwilym_M_

Jenkins_Gregory_C_Reinsel_and_Greta_M_Ljung_2015_Published_by_John_Wiley_and_Sons_Inc_Hoboken_New_Jersey_pp_712_ISBN_

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Bryman, A. (2012). *Social research methods* (4th ed.). Oxford University Press.

Civicus. (2023). Civicus monitor global findings africa [Accessed: 2024-07-19]. https://monitor.civicus.org/globalfindings/africa/

CNN. (n.d.). Cable news network [Retrieved n.d.].

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Crisis Group. (n.d.). Crisis watch [Accessed: 2024-07-19]. https://www.crisisgroup.org/crisiswatch

Dancy, G., & Fariss, C. J. (2017). The illusion of progress: Understanding the decline of international human rights reporting. *American Political Science Review*, *111*(3), 1–16. https://doi.org/10.1017/S0003055417000175

DefendDefenders. (2023). Bi-annual report: April 2023 - october 2023 [Retrieved March 22, 2024]. https://defenddefenders.org/bi-annual-report-april-2023-october-2023/

Earl, J. (2003). Tanks, tear gas, and taxes: Toward a theory of movement repression. *Sociological Theory*, *21*(1), 44–68. https://doi.org/10.1111/1467-9558.00174

Forst, M. M. (2018). World report on the situation of human rights defenders.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://www.researchgate.net/publication/13853244_Long_Short-term_Memory

Human Rights Watch. (2024). *South africa's obligations under international and domestic law* [Accessed: 2024-07-10]. Human Rights Watch. https://www.hrw.org/legacy/reports98/sareport/App1a.htm

International Service for Human Rights. (2024). Who are human rights defenders? [Accessed: 2024-03-22]. https://ishr.ch/about-human-rights/who-are-human-rights-defenders/

Marazzi, A. (2016). The data pipeline architecture. https://medium.com/the-data-experience/building-a-data-pipeline-from-scratch-32b712cfb1db

Medvedeva, M., Vols, M., & Wieling, M. (2020). Using machine learning to predict decisions of the european court of human rights. *Artificial Intelligence and Law*, *28*(2), 237–266. https://doi.org/10.1007/S10506-019-09255-Y/FIGURES/5

Missing Voices. (2024). Missing voices [Accessed: 2024-04-23]. https://www.missingvoices.or.ke/

Pandey, S. M., Agarwal, T., & Krishnan, N. C. (2018). Multi-task deep learning for predicting poverty from satellite images. *The Thirtieth AAAI Conference on Innovative Applications of Artificial Intelligence (IAAI-18)*, 7793–7798. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/viewFile/16441/16388

Quinn, J. A., Nyhan, M. M., Navarro, C., Coluccia, D., Bromley, L., & Luengo-Oroz, M. (2018). Humanitarian applications of machine learning with remote-sensing data: Review and case study in refugee settlement mapping. *Philosophical Transactions of the Royal Society*. https://doi.org/10.1098/RSTA.2017.0363

Ran, S., Lu, D., Tetreault, J., Cahill, A., & Jaimes, A. (2023). A new task and dataset on detecting attacks on human rights defenders [Accessed: 2024-03-22]. https://www.ohchr.org/Documents/

Reporters Without Borders. (2023). *Freedom of expression and civic space: Global overview* [Accessed: 2023-12-01]. Reporters Without Borders. https://rsf.org/en

Sampaio, N. (2024). A guide to digital and ai leadership in asia [Accessed: 2024-06-05]. https://www.imd.org/blog/digital-transformation/descriptive-analysis/

Smith, J. (2020). *Closing civic space: The global rise of government restrictions on civil society*. Oxford University Press.

Sudan Tribune. (n.d.). Sudan tribune - plural news and views on sudan [Accessed: 2024-07-10]. https://sudantribune.com/

U.S. Department of State. (2023). 2023 country reports on human rights practices [Retrieved March 22, 2024]. https://state.gov/reports/2023-country-reports-on-human-rights-practices/