Group 5: Conflict Intensity Prediction in Ukraine War

Supawit Sutthibolican Talika Gupta Yuechuan Zhang

Abstract

This project aims to predict conflict intensity in the Russian-Ukraine War by analyzing the ACLED (Armed Conflict Location and Event Data) and Oryx datasets. Using a comprehensive data pipeline, we first clean and enrich the data, and then extract relevant features from conflict event notes, such as attack types and equipment losses. We perform extensive data visualization to gain insights about the trends in the data. We use clustering to label conflict intensity for each record and subsequently use LSTM for time-series classification to predict future conflict intensity given previous data. We report macro-averaged F-1 score for our model performance.

1 Introduction

As the Russia-Ukraine War continues, significant efforts have been dedicated to promoting peace talks and providing humanitarian aid. In this project, we aim to detect peace windows with relatively low conflict intensity by forecasting war intensity based on the ACLED [1] and Oryx [2] dataset.

2 Motivation

The prolonged Russia-Ukraine War has created a grim reality for those living in war zones. Despite the ongoing conflict, certain periods with significantly lower conflict intensity could offer valuable opportunities for deploying humanitarian aid and fostering peace talks. In this project, we aim to detect such time windows and provide actionable insights to the global community.

3 Related Work

Existing work [3] primarily focuses on predicting the number of future attack events. The work by [5] first fact-checks the conflict dataset of the GDELT project, and then performs change point analysis to identify important trends in the data. There are also studies [4] that focus on predicting the number of fatalities in armed conflicts. However, we believe that differentiating between various attack and weapon types is necessary for accurately capturing the overall status of the war, which we aim to address in our work.

4 Dataset

Our analysis is based on the ACLED and Oryx datasets. The ACLED dataset contains conflict event reports from various sources such as news agencies and social media platforms. Each event includes attributes such as date, event type, actors involved, and location. For this project, we work with the ACLED Europe & Central Asia dataset and filter for events related to the Ukraine War. More specifically, we focus on five event types that are significantly related to military operations including air/drone strikes, armed clashes, shelling/artillery/missile attacks, government regaining territory, and disrupted weapon use. The Oryx dataset is a public Google Sheet updated based on the Oryx blog site which documents military equipment losses backed by photo evidence. Each record specifies the daily loss and capture number of Ukraine and Russian military equipment.

5 Methodology

5.1 Data Preprocessing

We notice that the note attribute of the ACLED dataset provides a text description of the attack for each event. For example, the note often contains information about the type of equipment used as well as the number of equipment lost. We believe that this information is essential for predicting the war intensity, and hence we extract them from the notes by mining for keywords of interest such as missile, drone, and air strike. The result is a vector containing the ACLED data along with the extracted features such as boolean flags on whether an air strike happened on a day and numerical values on equipment loss and firing count.

We also enrich our dataset with the Oryx dataset containing confirmed equipment loss

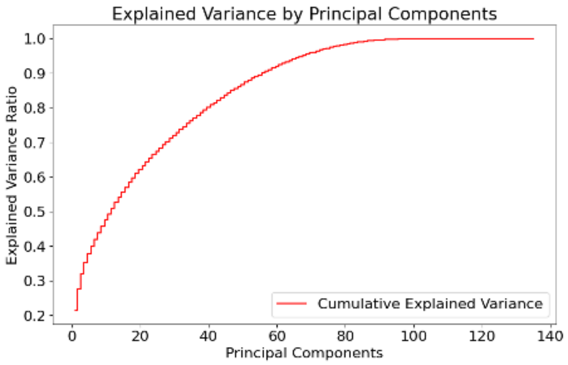

Figure 1: Principal Component Analysis on Preprocessed Dataset

numbers for each day (such as tanks, weapons, etc.). We believe that these numbers demonstrate war investment and are therefore good indicators of war intensity. We join each Oryx entry to the preprocessed ACLED records by date and the final product of our data preparation process is a dataset of 134-dimensional vectors where each vector contains information about a day during the war period between January 2018 and September 2024.

5.2 War Intensity Labeling through Clustering

Given the feature vector for each day during the war, we apply K-means clustering to identify war intensity labels that will serve as the target variable for our prediction task. Prior to performing clustering, We apply principal component analysis (PCA) to our feature vectors to help us understand the variance distribution of the data. As shown in Figure 1, 68 principal components are sufficient to capture 95% of the variance in the data.

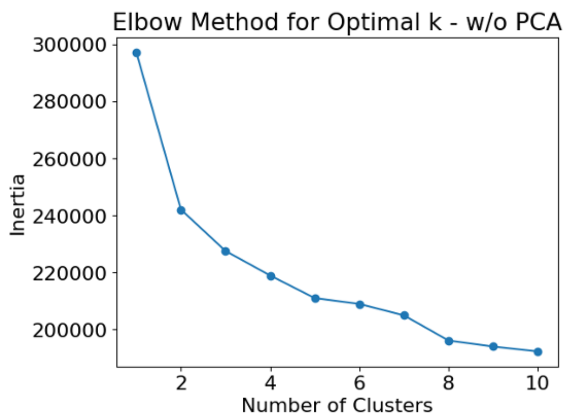

Figure 2: Elbow-chart on Original Dataset

Given the importance of feature interpretability, we investigate and compare the cluster sum

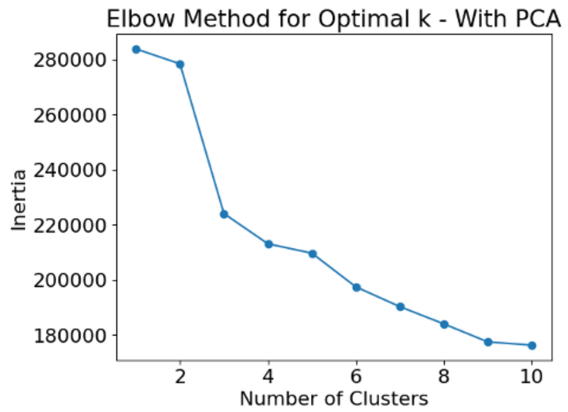

Figure 3: Elbow-chart on Feature-Reduced Dataset

of squares and Silhouette Scores after performing K-means clustering on both the original dataset and the principle-feature-reduced dataset to evaluate whether reducing the number of features improves the clustering outcome. Figure 2 and Figure 3 respectively show the elbow-chart for clustering on the original and feature-reduced dataset, which both indicate an elbow at $K = 3$. Furthermore, we observe a Silhouette Score of 0.40 for the feature-reduced dataset and a score of 0.38 for the original dataset which indicates similar clustering quality. Therefore, we proceed with the original feature version to maintain feature interpretability for our analysis.

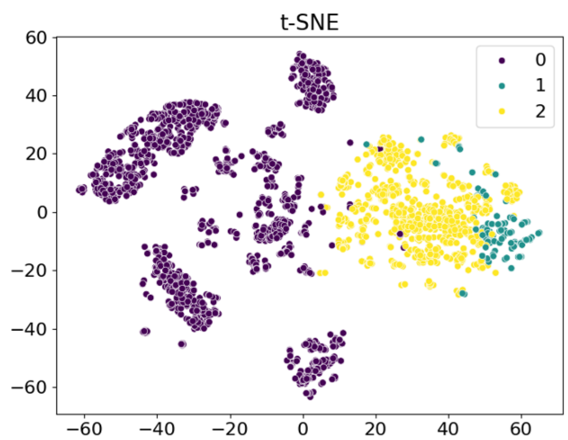

Figure 4: T-SNE Visualization of K-means clustering

We apply t-SNE and UMAP to our K-means clustering to visualize the clustering structure. The results, shown in Figure 4 and Figure 5, demonstrate the separation of data points into three distinct clusters. By investigating the cluster summary on fatalities and equipment losses, we interpret Cluster 0 as representing low intensity, Cluster 2 as medium intensity, and Cluster 1

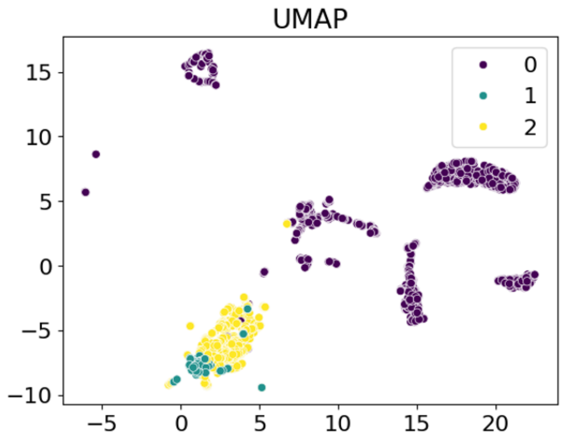

Figure 5: UMAP Visualization of K-means clustering

as high intensity. The result of labeling from K-means clustering is shown in Figure 6. As can be seen from the figure, there is a strong correlation between the clusters and time i.e. the 3 clusters formed are clearly separated through time. The “low” cluster appears uniformly from 2017 to 2021, while “medium” appears almost uniformly from 2022 to mid 2024, and towards the end only the “high” cluster appears. This correlates well with the war intensity periods that we observe in the real world. 2017 to early 2022, the invasion had not begun and therefore each day is mapped to a low cluster intensity. With the start of the invasion from February 2022, the conflict intensity picked up and it can be classified as a “medium” intensity level. Towards the end of 2024, we see an increase in the number of violent events in our dataset and thus ’high’ level is appropriate for the days in late 2024.

Aside from K-means clustering, we also attempt DBSCAN, Agglomerative Clustering with the Ward criterion, and Gaussian Mixture Models (GMM). The result of DBSCAN is suboptimal. The results of both Agglomerative Clustering and Gaussian Mixture Models successfully group the data into three clusters, similar to K-means. However, both methods exhibit inaccuracies, clustering some dates before the beginning of full-scale the war in 2022 as medium intensity as shown in Figure 7. We believe that the result of K-means clustering most accurately matches the public understanding of the progression of the war as explained in Section 6.1 and therefore we proceed with using it for data labeling.

5.3 Time series prediction - Training and Evaluation

Since the labels are so clearly separated in different time periods, it is difficult to split the data in a way that ensures all classes are represented well in the train and test set both, while not shuffling the temporal order in which the samples (each sample is a sequence of 90 days) appear. We split our dataset for training and testing such that the first 10%, the middle 6%, and the last 4% of data will be used for testing, while all prior data to it will be used for training. This ensures all class labels are represented in the test and training sets. We use a sliding window approach to create training samples, with a sequence length of 90 days. The sequence length of 90 days was chosen to ensure the model captures adequate historical/seasonal trends. Each training/test sample will consist of 90 days, and the model will predict the conflict intensity for the 91^{th} day. This is compared with the actual conflict intensity label on that day.

We select an LSTM because of its capability to model sequential data and capture temporal dependencies effectively, which are essential for forecasting conflict intensity based on historical trends. We train the LSTM using the prepared data for 20 epochs. Before training, the features are scaled using Min-Max scaling to ensure they are within a range suitable for LSTM processing. We use the sparse categorical cross-entropy loss and the Adam optimizer for training. Since our dataset is highly class-imbalanced, we assign weight values to the class labels. We use this to weight the loss function, giving more weightage to samples from underrepresented classes. The LSTM has 2 layers, with each layer containing 50 units and a dropout rate of 20% to prevent overfitting. We use the accuracy metric for evaluation which counts the number of predictions that match with the actual conflict intensity label. After training, we arrive at an accuracy of 92.4% on the test set. The per-class and average F1 scores are shown in Table 1. The highest F1 score we achieve is 0.980 for the high class, and the lowest is 0.736 for the medium class. We achieve a fairly good macro-averaged F1 score of 0.888 and a micro-average F1 score of 0.924. The confusion matrix is shown in Figure 8.

5.4 Data Visualization

We perform data visualization to understand the trends and distributions of our dataset. We first visualize the trend of the conflicts across months

Figure 6: War Intensity Clusters Derived from Applying K-Means with a Cluster Count of Three

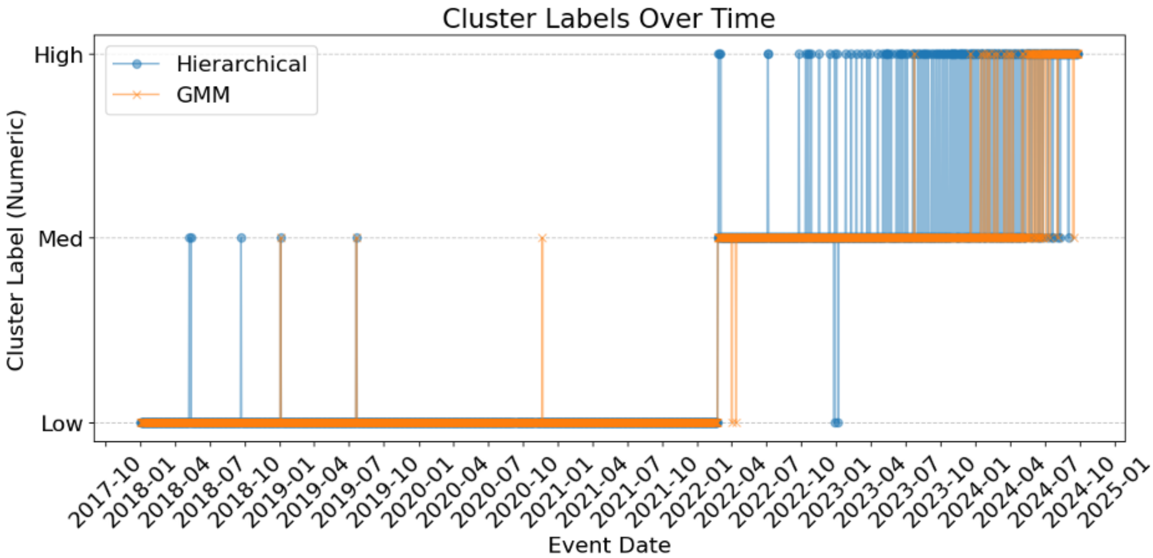

Figure 7: Hierarchical VS GMM Clustering Results

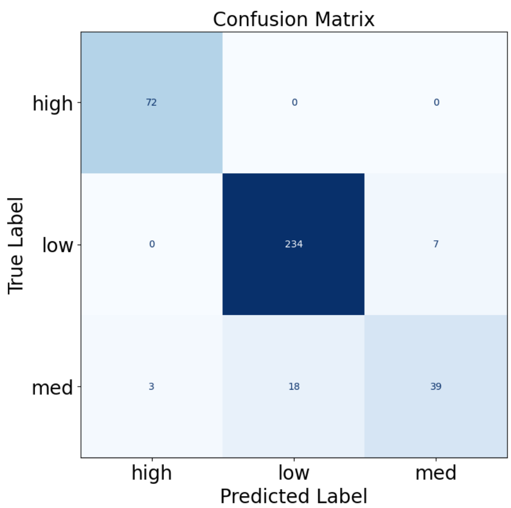

Figure 8: Confusion matrix

Label	**F1 Score**
High	0.980
Low	0.949
Med	0.736
Macro-Averaged F1 Score	0.888
Micro-Averaged F1 Score	0.924
Weighted-Averaged F1 Score	0.920

Table 1: Per-Class F1 Scores and Averaged F1 Metrics

from 2018 to 2024 using a heatmap in Figure 9. We can see a sharp increase in conflict events starting in 2022, which marks the start of Russia's full-scale invasion of Ukraine. We can also see a seasonal trend, with peaks in June-September while winter months have relatively fewer conflicts. There is a noticeable dip during 2020-2021, which could be attributed to COVID-19.

The line graphs in Figure 10 show the temporal trends for each event type. We observe a sharp increase in shelling/artillery/missile attacks in 2022, which began to decline mid 2024. Also, there has been a consistent upward trend in air/drone strikes since the full-scale invasion began, which is continuing through the present. Thus, from these figures, we can observe a clear temporal trend in the data which can be exploited for prediction.

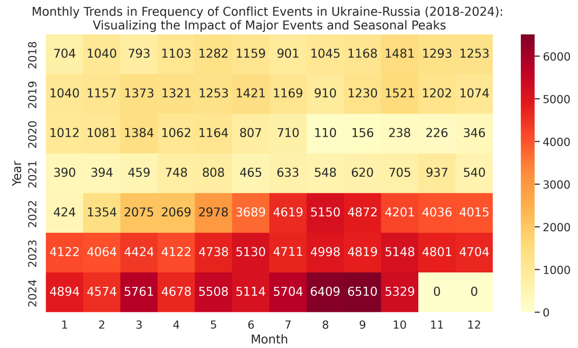

Figure 9: Monthly Trends in Frequency of Conflict Events in Ukraine-Russia (2018-2024)

6 Results

6.1 Cluster Interpretation

Figure 6 shows the war intensity label for each day corresponding to its K-means cluster. The result demonstrates that K-means clustering is able to successfully identify major phases in the Russo-Ukrainian War according to the Wikipedia timeline [6]. The low-intensity period before February 2022 corresponds to the initial military unrest in the Donbas region. The war intensity rose to medium after Russia declared full-scale war against Ukraine in February 2022. The scatter drop of intensity starting in November 2022 matches the general understanding of a military stalemate. Finally, the high-intensity period since the beginning of 2024 signals a major elevation in war investment following the 2023 winter campaigns where both sides have increased their reliance on drones and missiles.

6.2 Prediction Evaluation

In terms of time series prediction outcomes, we achieve a good average F1 score as can be seen from Table 1. However, the model performs poorly for the medium class - the F1 score for the medium class is the lowest. This underperformance may stem from overlapping days in sequences with the low and high classes, making it challenging for the model to distinguish between them. Future work could involve feature engineering to identify additional distinguishing characteristics for this class. Also exploring different class balancing techniques or collecting more data could enhance performance for the medium class, leading to a more balanced and accurate model.

7 Conclusion

In this project, we have investigated and visualized trends in conflict intensity of the Russo-Ukraine War and performed intensity prediction by 1) extracting features from the ACLED and Oryx dataset 2) identifying intensity labels from clustering 3) forecasting intensity by leveraging LSTM. Our three-class intensity clustering matches the common understanding of the development of the war and our LSTM-based conflict intensity prediction performs fairly well. While war intensity forecasting for the purpose of civilian evacuation and humanitarian aid remains an open challenge as it necessarily involves a more complex prediction of geographic location, our project has demonstrated the potential of macro-level war intensity prediction through open-source datasets.

Figure 10: Trends in Conflict Event Types in the Ukraine-Russia Conflict (2018-2024)

References

[1] ACLED (2024). *Armed Conflict Location & Event Data Project (ACLED) Dataset* (Version: 2024-09-27). Available at: `https://acleddata.com`

[2] Oryx. *Attack On Europe: Documenting Russian Equipment Losses During The Russian Invasion Of Ukraine*. Available at: `https://www.oryxspioenkop.com/2022/02/attack-on-europe-documenting-equipment.html`

[3] ACLED. *ACLED Conflict Alert System (CAST)*. Available at: `https://acleddata.com/conflict-alert-system/`

[4] Håvard Hegre, Paola Vesco, Michael Colaresi, et al. *The 2023/24 VIEWS Prediction Challenge: Predicting the Number of Fatalities in Armed Conflict, with Uncertainty*. Available at: `https://arxiv.org/pdf/2407.11045`

[5] Swetha Keertipati, Bastin Tony Roy Savarimuthu, Maryam Purvis, and Martin Purvis. *Multi-level Analysis of Peace and Conflict Data in GDELT*. Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis, pp. 33–40, 2014. Association for Computing Machinery, New York, NY, USA. Available at: `https://doi.org/10.1145/2689746.2689750`

[6] Wikipedia. *Timeline of the Russian invasion of Ukraine*. Available at: `https://en.wikipedia.org/wiki/Timeline_of_the_Russian_invasion_of_Ukraine`