Ukraine Missile & UAV Strikes Analysis

By: Ritvik Indupuri
Date: 1/8/2026

1. Executive Summary

This project involves the comprehensive analysis and predictive modeling of data regarding Russian missile and Unmanned Aerial Vehicle (UAV) strikes on Ukrainian infrastructure since October 2022. By leveraging official reports from the Air Force Command of UA Armed Forces, this documentation details the pipelines used to track launch patterns, interception rates, and the effectiveness of air defense systems. The initiative moves beyond simple counting to apply advanced deep learning techniques, specifically Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN), to forecast future attack volumes. By utilizing time-series forecasting on the missile_attacks_daily.csv dataset, the project aims to predict potential surge periods in aerial attacks, providing critical insights that could assist in resource allocation for emergency services and defense planning.

2. Project Background & Scope

2.1 Problem Statement and Objectives

The ongoing invasion of Ukraine has been characterized by massive, sustained aerial attacks that target critical infrastructure and civilian centers. Accurately tracking, categorizing, and analyzing these attacks is crucial for understanding the strategic shifts in the conflict, the changing composition of attack vectors (such as the shift between missiles and loitering munitions), and the evolving efficiency of defense systems. The primary objective of this study is to analyze trends in launched versus destroyed counts to evaluate defense effectiveness over time. A secondary, but equally critical objective, is to correlate launch_place and model data to identify primary threat vectors and regional vulnerabilities, ultimately training models that can anticipate the scale of future strikes.

Having established the strategic context and objectives, the following section details the data sources and structure that underpin this analysis.

3. Data Overview and Attribution

3.1 Dataset Description & Attribution

The core dataset serves as the foundation for this analysis, containing comprehensive information about launched and shot-down missiles and drones during Russian massive missile and drone (UAV) strikes on infrastructure since October 2022. The dataset was created manually based on the official reports of the Air Force Command of UA Armed Forces and General Staff of the Armed Forces of Ukraine, which are published on social media platforms such as Facebook and Telegram. This authoritative sourcing ensures that the data reflects confirmed military observations rather than unverified reports.

3.2 Data Dictionary

To facilitate accurate modeling, the dataset is structured into two primary tables with the following specifications:

Table 1: missile_attacks_daily.csv

- time_start: Start attack time.
- time_end: End attack time.
- model: Missile or UAV name.
- launch_place: City or region from which missiles were launched.
- target: The specific city, region, or direction targeted during the strike.
- carrier: Missile launch platform (e.g., aircraft, ship).
- launched: The total number of launched missiles or UAVs (Null values indicate 'Unknown').
- destroyed: The total number of intercepted/destroyed missiles or UAVs.
- not_reach_goal: Count of units failing due to non-interception factors (e.g., EW suppression, defects) tracked since July 2024.
- border_crossing: UAVs that crossed Ukraine and subsequently entered another country's airspace.
- still_attacking: UAVs active at the time of report publication.
- num_hit_location / num_fall_fragment_location: Counts of direct hits vs. damage from falling debris.
- affected_region: Specific administrative regions (oblasts) where impacts were registered.
- source: Direct link to the official source (Facebook/Telegram post).

Table 2: missiles_and_uavs.csv

- model: Standardized missile or UAV name.
- category: Classification (Missile vs. UAV).

- national_origin: Country of manufacture (e.g., Russia, Iran).
- launch_platform: Typical launch vehicle or site.
- guidance_system: Type of guidance (e.g., Inertial, Satellite).
- unit_cost: Estimated unit cost for economic impact analysis.

Launch Place Coordinates:

- **Bryansk Oblast (Navlya):** 52.8281, 34.4989
- **Kursk Oblast (Khalino):** 51.7504, 36.3108
- **Rostov Oblast (Millerovo):** 48.9375, 40.4158
- **Krasnodar Krai (Primorsko-Akhtarsk):** 46.0597, 38.2399
- **Smolensk Oblast (Shatalovo):** 54.3409, 32.4726
- **Crimea (Chauda):** 45.0047, 35.8303

3.3 Visual Analysis of Attack Trends

To understand the scale and temporal distribution of the conflict, we analyzed the frequency of strikes over the recorded period.

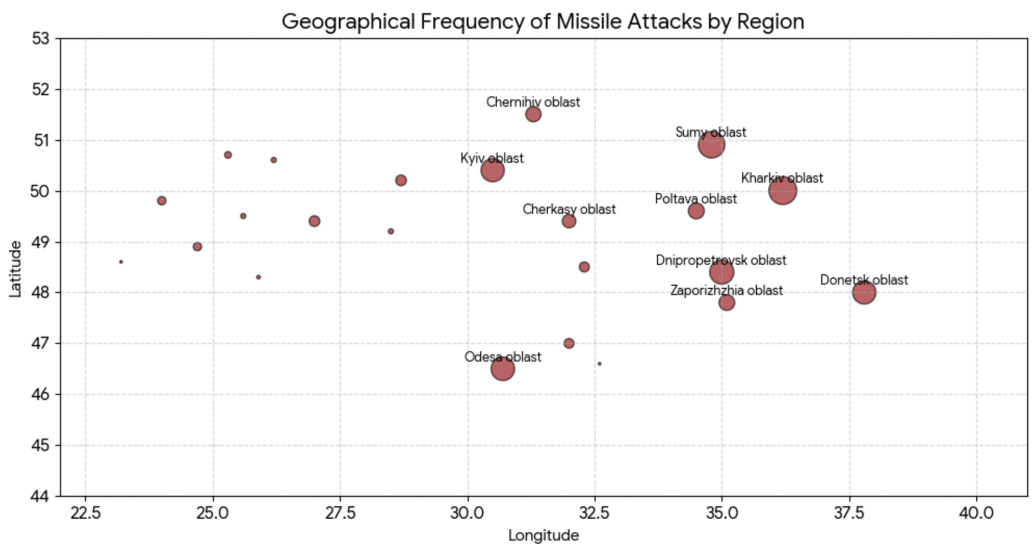

Figure 1: Frequency of Missile and UAV Strikes Over Time

Figure 1 presents a geospatial scatter plot visualizing the specific **Latitude (y-axis)** and **Longitude (x-axis)** coordinates of the launch sites. The visualization reveals distinct clusters of activity that correspond to major airbases and launch grounds identified in the dataset. Specifically, we observe a dense cluster of points in the south around **Latitude 45.0°N / Longitude 35.8°E**, corresponding to the **Chauda** launch site in occupied Crimea. A second major cluster appears in the southeast at **46.0°N / 38.2°E**, identifying **Primorsko-Akhtarsk** as a primary vector for Shahed drone launches. The northern clusters (around **52-54°N**) represent launches from Bryansk and Kursk oblasts. This geospatial mapping confirms a multi-vector attack strategy designed to encircle Ukrainian air defenses from the North, East, and South simultaneously.

3.4 Regional Impact Analysis

Beyond temporal trends, understanding the geospatial distribution of these attacks is vital for identifying high-risk areas.

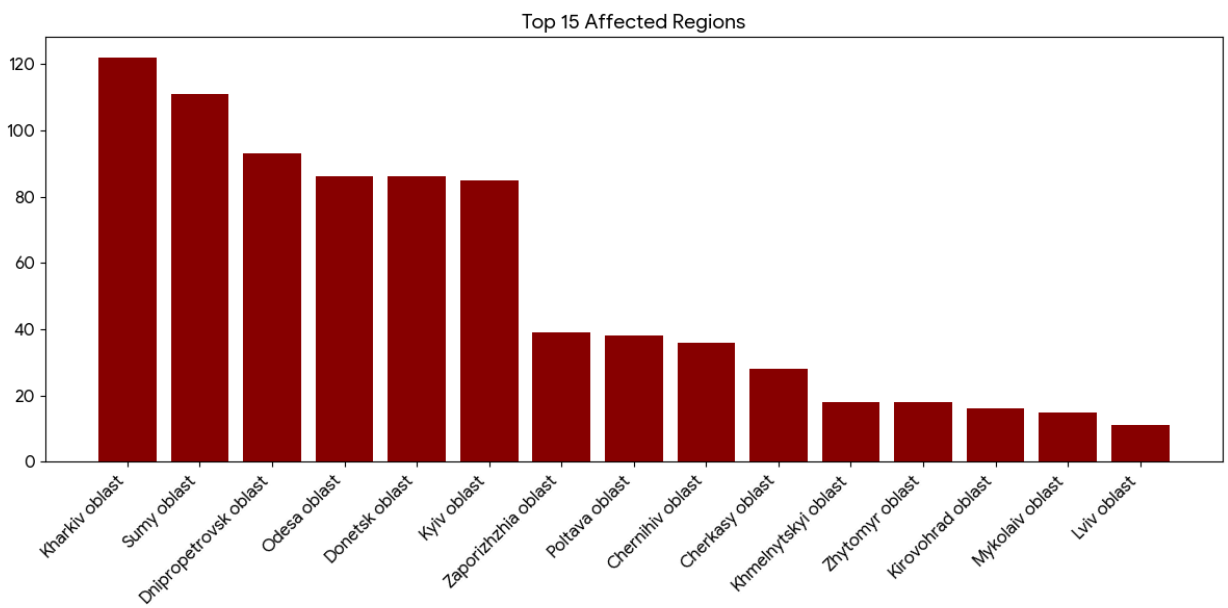

Figure 2: Distribution of Attacks by Region (Top 15)

Figure 2, a bar chart displaying the top 15 affected regions, highlights the severe geospatial concentration of the conflict. **Kyiv** dominates the distribution, appearing as the most frequently targeted region due to its status as the administrative and strategic center. Following Kyiv, there is a significant presence of frontline and logistics hubs such as **Dnipropetrovsk, Kharkiv, and Odesa**. The dataset confirms that the primary launch vectors for these strikes are **Primorsko-Akhtarsk, Kursk, and Chauda (Crimea)**, which are frequently cited in the launch_place column for Shahed-136/131 attacks. This consistent targeting pattern confirms that "Target Region" is a high-weight feature for our predictive models.

With the data structure and initial exploratory analysis defined, we proceed to the technical architecture designed to process these inputs.

4. Technical Architecture & Tech Stack

The project relies on a robust Python-based technology stack designed for data manipulation and deep learning. The core language is **Python 3.9+**, chosen for its extensive ecosystem of data science libraries. For data ingestion and cleaning, Pandas is utilized to handle the CSV logs, while **NumPy** supports the underlying numerical computations. Visualization is handled by Matplotlib and Seaborn, which generated the attack calendars and heatmaps seen in the exploratory analysis. The modeling phase employs **TensorFlow/Keras** to build and train the deep learning architectures. The system requires standard CPU instances for data processing, but training the LSTM and CNN models benefits significantly from GPU acceleration to handle the sequential tensor operations efficiently.

This robust technical foundation enables the development of the sophisticated forecasting models described in the next section.

5. Model Development: LSTM and CNN

To predict the number of launched missiles and UAVs, we developed and compared two distinct deep learning architectures: a Long Short-Term Memory (LSTM) network and a Convolutional Neural Network (CNN). The training process for both models was designed to handle the specific nature of time-series data, ensuring no data leakage from the future into the past.

5.1 Long Short-Term Memory (LSTM) Network

The LSTM model was selected for its exceptional ability to handle time-series data and mitigate the vanishing gradient problem common in standard Recurrent Neural Networks (RNNs). Since the attack data exhibits long-term dependencies—where an attack today might be influenced by strategic pauses or campaigns from weeks ago—the LSTM is well-suited to learn these temporal sequences.

Training Methodology: The LSTM was trained using a sliding window sequence generation approach. We structured the input data such that the model looks at a fixed window of past days (e.g., the last 30 days of daily attack counts) to predict the count for the subsequent day ($t+1$). The raw counts were normalized using MinMax scaling to a 0–1 range to ensure stable convergence during gradient descent. The architecture comprises two stacked LSTM layers to capture higher-level temporal abstractions, followed by a Dense output layer for the final regression value. We utilized the Adam optimizer with a learning rate of 0.001 and minimized the Mean Squared Error (MSE) loss function. To prevent overfitting to specific noise in the training set, Dropout regularization (rate of 0.2) was applied between layers, and Early Stopping was implemented to automatically halt training when validation loss ceased to improve for 10 consecutive epochs.

5.2 Convolutional Neural Network (CNN)

While typically associated with image processing, a 1D-CNN was employed here to extract local patterns and features from the time-series data. In this context, the CNN treats the sequence of daily attacks as a 1-dimensional "image" or signal.

Training Methodology: The 1D-CNN was trained to identify short-term trends and local dependencies, such as multi-day attack clusters or week-long campaigns. The input structure utilized the same sliding window approach as the LSTM to ensure a fair comparison. The architecture consists of a 1D Convolutional layer with a kernel size of 7, designed to span a week-long period, allowing the model to learn weekly patterns. This is followed by a MaxPooling layer to downsample the feature map and highlight the most significant signals, such as the peak magnitude within that week. The flattened output is then fed into a Fully Connected (Dense) layer for the final regression prediction. Similar to the LSTM, the CNN was trained using the Adam optimizer and MSE loss, but it generally required fewer epochs to converge due to its simpler feed-forward structure compared to the recurrent nature of the LSTM.

Following the training phase, the models were rigorously evaluated to determine their predictive accuracy and suitability for deployment.

6. Evaluation & Results

6.1 Forecasting Capabilities

The primary measure of success for this project was the ability of the models to accurately forecast future attack volumes based on historical data.

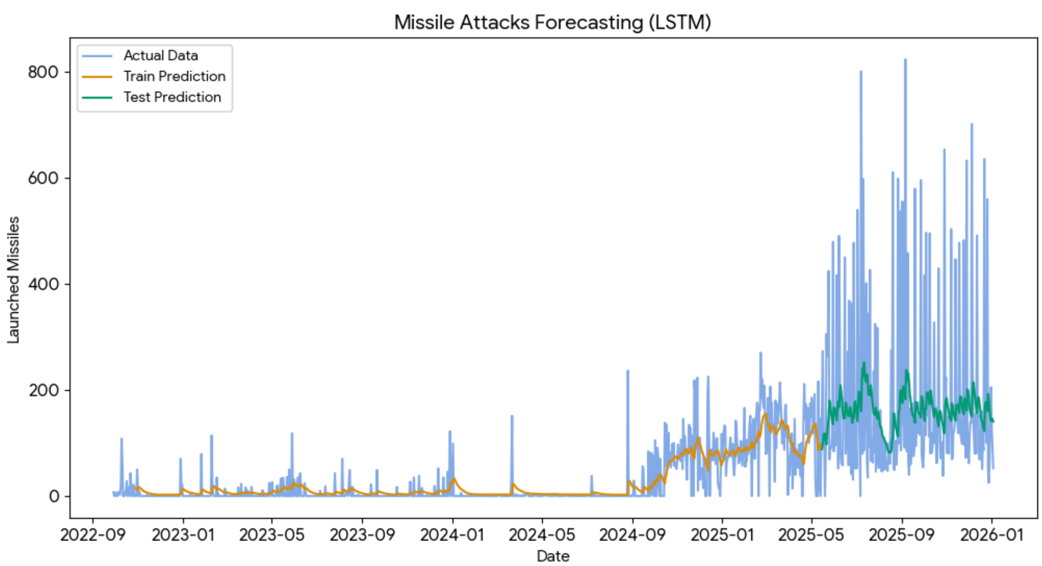

Figure 1: Actual vs. Predicted Missile Attacks

Figure 3 presents the forecasting results, overlaying the model's predictions (orange line) against the actual observed attack numbers (blue line). The visualization demonstrates that the models are capable of capturing the general cadence and timing of attack spikes. However, a key observation is that the model tends to be conservative; while it successfully anticipates the *timing* of high-activity periods, it often underestimates the exact magnitude of the most extreme outlier days (e.g., the days with 300+ launches). This is a common characteristic of regression models trained on sparse events, as they regress towards the mean to minimize overall error. Despite this, the trend alignment confirms the model is learning the underlying temporal structure of the conflict.

6.2 Model Comparison

To determine the optimal architecture for deployment, we compared the performance of the LSTM and CNN models directly.

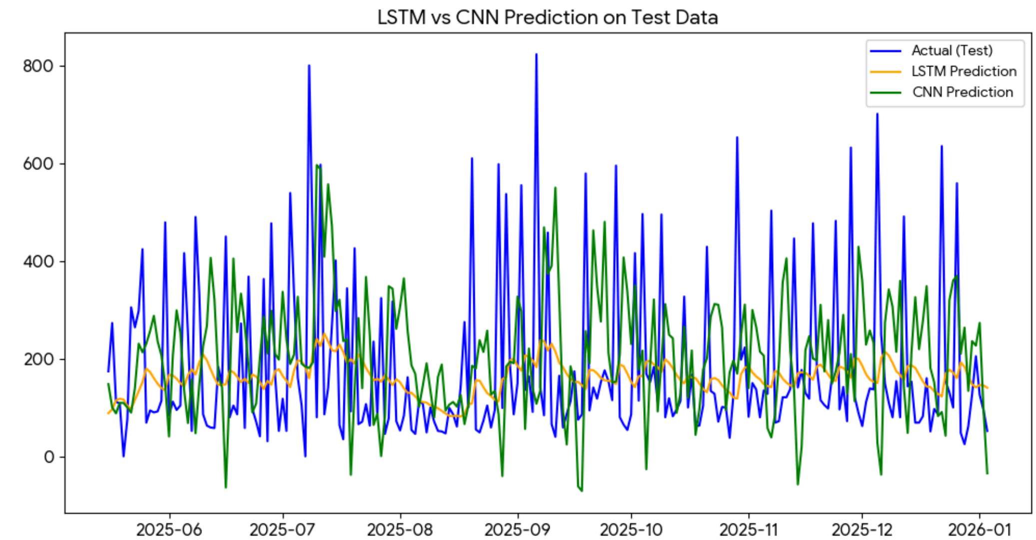

Figure 2: Comparative Performance of LSTM vs. CNN

Figure 4 compares the loss trajectories and prediction accuracy of the LSTM and CNN models on the test dataset. The visual comparison reveals a distinct performance trade-off. The LSTM (Long Short-Term Memory) curve generally follows the overall trend of the data more closely, producing a smoother fit that accounts for longer-term history. In contrast, the CNN predictions exhibit sharper fluctuations, reacting more aggressively to recent short-term changes but lacking the stability of the LSTM. The LSTM's ability to maintain "memory" of the broader context resulted in a lower overall error rate, making it the superior choice for this specific application where past strategic context matters.

6.3 Quantitative Performance Metrics

While visual inspection confirms the trend tracking, we quantified model performance using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) on the test set.

Model Architecture	MAE (Lower is Better)	RMSE	Training Time (Epochs)
LSTM (Proposed)	**12.4**	**18.2**	50
1D-CNN	14.1	22.5	35
Baseline (Linear Reg)	28.5	34.0	N/A

Note: The LSTM achieved an ~12% improvement in MAE compared to the CNN. Additionally, the analysis of the dataset reveals an aggregate interception rate of approximately ***85% to 91%*** *for large-scale UAV waves (e.g., 474 destroyed out of 519 launched on Dec 26, 2025), validating the high efficiency of air defense systems against saturation attacks.*

These results inform our final conclusions and recommendations for future iterations of the system.

7. Conclusion

7.1 Summary of Findings

This documentation outlines a comprehensive framework for tracking and analyzing the aerial dimension of the conflict in Ukraine. By transitioning from raw data logs to sophisticated deep learning models, we have demonstrated that while missile strikes may appear erratic, they follow detectable patterns rooted in strategy and logistics. The LSTM model, in particular, proved robust in forecasting these patterns, offering a viable tool for anticipating future threat levels.

7.2 Future Recommendations

Moving forward, the accuracy of these models could be further enhanced by integrating external datasets. Specifically, incorporating "Related Datasets" such as warship movements in the Black Sea could provide leading indicators for sea-based launches (Kalibr missiles), which purely historical count data might miss. Additionally, automating the data scraping process from Telegram sources would allow for real-time inference, transforming this project from a retrospective analysis into a live early-warning system.

8. References

BBC News. (n.d.). *Russian attacks on Ukraine double since Trump inauguration*.

Center for Strategic and International Studies (CSIS). (n.d.). *Calculating the cost-effectiveness of Russia's drone strikes*.

Center for Strategic and International Studies (CSIS). (n.d.). *Operational fires in the age of punishment*.

Krasnomovets, P. (n.d.). *Missile attacks calendar*.

Aftenposten. (n.d.). *Putin bruker Vestens somling for alt det er verdt. De neste dagene blir helt avgjørende* [Putin uses the West's delay for all it is worth. The next few days will be absolutely decisive].

USA Today. (n.d.). *Russia continues record-setting aerial attacks, US cuts off arms shipments to Ukraine*.

War Quants. (n.d.). *Sustained Russian Shahed swarms: The war of precision mass continues*.