Assessing the trustworthiness and real-world feasibility of machine learning models for conflict forecasting

SUPERVISOR: MARIJN ROELVINK

RESPONSIBLE PROFESSOR: DR. CYNTHIA LIEM

AUTHOR: ALEXIA-IUSTINA GAVRILĂ

A.I.GAVRILA-1@STUDENT.TUDELFT.NL

TU Delft

1. INTRODUCTION

Growing need for humanitarian assistance

Decreasing financial resources

⟶ The solution? **Early-action!**

- **What it is:** Early-action refers to proactive interventions based on forecasts, aiming to mitigate crises before they fully unfold.
- **Why it matters:** It is more cost-effective, can reduce suffering, and allows better planning in settings with limited resources.
- **How it's done:** ML models forecast where crises may escalate, helping humanitarian actors prioritze efforts and respond faster.

Focus area: **Armed Conflict Forecasting**

2. RESEARCH GAPS

- Focus on predictor variables and conditioning factors in past research.
- Lack of focus on model validation methods and operational applicability.
- Limited attention to trustworthiness and real-world feasibility of ML systems.
- Absence of evaluation for use in practical humanitarian settings.

3. RESEARCH QUESTION

How reliable and feasible are machine learning systems for conflict forecasting in real-world humanitarian contexts?

RQ1: How trustworthy and accurate are existing ML models for conflict forecasting, based on their reported performance metrics and validation practices?

RQ2: Under what contextual conditions are these models practically deployable for real-world humanitarian decision making?

4. METHODOLOGY

This study follows the **SALSA** strategy to conduct a systematic review of literature on conflict forecasting using ML.

SEARCH

Search query: designed to intersect conflict-related terms with those related to ML and early warning.

⟶ **Result:** 494 studies across 3 databases (Scopus, IEEE, Web of Science)

APPRAISAL

Inclusion: English, peer-reviewed or recognized gray literature, with armed conflict and ML focus

Exclusion: Non-English, irrelevant fields, no DOI, no full-text

Result: 32 included studies

SYNTHESIS

Thematic categories:
- Forecasting scope and purpose
- Data sources and quality
- Modelling approaches
- Reliability and robustness
- Ethics and practical application

ANALYSIS

Process: cross-comparative evaluation of the 32 included studies.

Steps:
1. Comparing findings across studies within each category.
2. Identifying recurring trends and limitations.
3. Extract insights to address the research questions.

5. RESULTS

Forecasting scope and purpose:
- Forecast targets vary. 14 out of 32 studies forecast conflict occurrence. Others target conflict onset, type, and the number of fatalities.
- Most models focus on early warning (18 studies) and improvement of methodology (10 studies)

Data sources and quality
- Most used datasets: ACLED and UCDP.
- Data issues like missing values and class imbalance are common, especially in conflict-vulnerable regions.

Modelling approaches
- Most popular method: Random Forest (21 studies)
- Validation methods vary: cross-validation is the most common, used by 23 studies, but lacks transparency.

Reliability and robustness
- Most studies evaluate robustness through cross-validation but rarely report uncertainty explicitly.
- Only 8 studies offer detailed error analysis.

Ethics and practical application
- Only 5 studies discuss ethical aspects.
- Only 5 studies are in operational use.

6. CONCLUSIONS

Trustworthiness?

Most models show promising accuracy, **but**:
- Evaluation practices lack standardization.
- Uncertainty is rarely quantified.
- Error analysis is often missing.

⟶ **Trust is limited without transparency.**

Feasibility?

ML models are practically deployable **only when**:
- High-quality, timely, and disaggregated data is available.
- The model is interpretable for non-technical users.
- Uncertainty is clearly communicated.
- Ethical risks are acknowledged and mitigated.

⟶ **These conditions are rarely met, which limits real-world use.**

⟶ TAKEAWAY

- ML systems for conflict forecasting show strong technical potential but are **not yet** sufficiently reliable or feasible for widespread use in humanitarian contexts.
- They are held back by inconsistent evaluation, poor uncertainty communication, and low interpretability.
- To move beyond academic use, models must focus on usability, transparency, and ethics.