

Data Collection and Preprocessing Phase

Date	18 July 2024
Team ID	SWTID1720012105
Project Title	WarLens: Transfer Learning for Event Classification in Conflict Zones
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification

WarLens aims to gather and curate multimedia data (images and videos) from conflict zones to train and validate transfer learning models for event classification. The project focuses on high-risk areas identified through conflict maps, news reports, and satellite data. Key data sources include open-source intelligence (OSINT), social media platforms (Twitter, YouTube, Instagram), news agencies, NGOs, and satellite imagery providers. The data types include images and videos collected via web scraping tools from social media, news sites, and satellite imagery providers.

For WarLens, raw data will be sourced from diverse platforms. Images and videos will be gathered from social media platforms (Twitter, YouTube, Instagram), reputable news agencies, and satellite imagery providers. Additional sources include OSINT reports and NGO databases that provide real-time updates on conflict zones. The curated data will be systematically stored, ensuring data integrity and facilitating effective training of transfer learning models for accurate event classification in conflict zones.

Data Collection Plan

Section	Description
Project Overview	WarLens is an innovative machine learning project that utilizes transfer learning techniques to classify events in conflict zones by analyzing multimedia data such as images and videos. The project's objective is to enhance situational awareness and provide accurate event classification in high-risk areas.
Data Collection Plan	For WarLens, data will be collected from various sources, including social media platforms (Twitter, YouTube, Instagram), reputable news agencies, satellite imagery providers, and open source intelligence (OSINT) reports. This diverse data set will provide a comprehensive foundation for training transfer learning models to accurately classify events in conflict zones.
Raw Data Sources Identified	For WarLens, raw data will be sourced from social media platforms like Twitter, YouTube, and Instagram, which provide real-time user-generated content. Additional sources include reputable news agencies for verified multimedia reports, satellite imagery providers for detailed overhead views, OSINT reports for comprehensive conflict data, and NGOs for reliable on-the-ground information.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset	The data for WarLens consists of images and videos from social media platforms, verified multimedia reports from news agencies, detailed satellite imagery, comprehensive OSINT reports, and reliable on-the-ground content from NGOs.	https://drive.google.com/file/d/1_qiE733RgeD5f81AMal6scE_u2s4Tjza/view?usp=drivesdk	Image	84 MB	Private (with access)