جامعة أم القرى
UMM AL-QURA UNIVERSITY

Multi-hazard mapping using ML

Student Name	Student ID
Remas Saleh Ghandorah	445006854
Shatha Ali Alhumaidi	445000176
Haden Hassan Alshaar	445013952
Hadeel Mohammad Turki	445006064
Linah Ebrahim ALsayed	445002548

Section Number : 1

Instructor : Dr. Afaf Almehmadi

1.Abstract

2. Introduction

Disaster management is a critically important field, and the humanitarian aspect-addressing events like hurricanes, earthquakes, and wildfires-presents an exceedingly complex challenge that demands precise planning and maximum speed and accuracy in response. The stakes are immense, as human lives are at risk and resources are stretched to the limit.

The immediate response phase following a disaster is the most critical, where relief teams face colossal challenges due to the lack of accurate information regarding the extent of damage and the precise locations of victim gatherings.
It is well-established that delays or failures in swiftly and effectively identifying the whereabouts of those in need lead to late delivery of essential aid, exacerbating the suffering of those affected and hindering recovery efforts. Fortunately, recent advancements in Artificial Intelligence (AI) have fundamentally transformed how societies prepare for and respond to catastrophes. In this context, Machine Learning (ML) offers innovative and powerful solutions.

ML models have the capability to analyze vast amounts of unconventional and unstructured data (such as satellite imagery, remote sensing data, and streaming information from social media) in near real-time.

These techniques are adept at detecting complex patterns and providing reliable predictions for damage sites and human concentration points, offering crucial insights for decision-makers.

The objective of this research is to design a Conceptual ML Module capable of automatically identifying and mapping the gathering locations of disaster-affected populations.

The scope focuses on processing geo-spatial data to generate an initial map of critical affected sites requiring urgent relief, thereby enhancing the overall performance of disaster management systems and ensuring that aid reaches everyone.

3. Problem definition

- **Brief description of the chosen system or problem:**

  The system under study seeks to identify and map disaster-affected areas by analyzing pre- and post-disaster satellite imagery. These images often reveal changes to the physical environment such as collapsed structures, flooded zones, burned land, or blocked roads yet interpreting these changes manually is slow, labor-intensive, and not scalable during emergencies. The challenge lies in developing an automated method capable of detecting these spatial changes accurately and transforming them into actionable damage maps.

- **Motivation (Why this system needs improvement or how ML can help).**

  The information lacking is the main challenge facing first responders during an immediate disaster. Although there are traditional disaster management techniques, the limitations that make a more intelligent system needed and overcome the weaknesses of current systems, an ML system is required.

  ML combines the various data to produce a single, full operational picture that is valuable. Human analysts are unable to process the huge amount of images. Millions of data can be received and analyzed by ML to find patterns, categorize data, and detect important indications.

  Also, a dynamic map that updates immediately can replace a static map or a long-term damage assessment.With this dynamic map, first responders can modify their tactics as circumstances change and focus where they may be able to save more lives.

- **Early idea of what kind of ML approach might be suitable.**

  Given the difficulty of identifying spatial changes in satellite imagery and categorizing areas under time constraints, a tree-based ensemble model like Random Forest is a strong contender. It handles high dimensions and noise, which is in pre- and post-disaster satellite images. The model can learn the patterns that separate normal regions from damaged ones without requiring heavy preprocessing or complex feature.This makes it suitable for building an initial automated mapping module that can quickly flag critical zones and support real time decision-making during disaster response.

4. Literature Review

(1) A Hybrid Machine Learning Pipeline for Disaster Event Mapping from Social Media

Recent research highlights the growing importance of social media particularly Twitter as a real-time data source for disaster situational awareness. During crises, users frequently post information about infrastructure damage, flooding, injuries, and urgent needs. As a result, machine learning (ML) and natural language processing (NLP) techniques have been widely explored to automatically extract and map disaster-related information. However, earlier approaches faced persistent challenges, including noisy and unstructured text, limited geolocation accuracy, poor classification of humanitarian categories, and difficulty verifying the credibility of user-generated reports.

1. **Existing Methods and Techniques**

Early disaster-related tweet classification relied on traditional ML models such as Logistic Regression and Support Vector Machines (SVM) using TF-IDF features, as well as deep learning models including CNNs, LSTMs, and GRUs. While these methods achieved moderate success, they suffered from weak contextual understanding, limited generalization across different disaster events, and inconsistent performance on multi-class humanitarian labels. In contrast, recent studies demonstrate that transformer-based models—particularly BERT—significantly outperform earlier techniques by capturing richer semantic context, handling imbalanced datasets more effectively, and improving classification accuracy.

2. **Hybrid and Advanced Approaches**

Recent literature proposes hybrid ML pipelines that integrate multiple complementary techniques. These pipelines typically combine Named Entity Recognition (NER) to extract place names, geocoding services (such as Google Maps APIs) to convert textual locations into coordinates, and fine-tuned BERT models for accurate humanitarian classification. Additionally, graph-based clustering has emerged as a powerful method for identifying repeated reports, filtering isolated or unreliable tweets, and grouping messages that describe the same real-world event. Compared to standalone deep learning or rule-based systems, hybrid models achieve higher spatial precision, improved credibility detection, and better scalability.

3. **Datasets and Evaluation**

Most studies rely on large-scale Twitter datasets collected via the Twitter API, often focusing on major disasters such as Hurricanes Harvey, Irma, and Maria. Datasets are typically cleaned by removing retweets, replies, and irrelevant content. Large datasets—sometimes exceeding tens of millions of tweets—enable robust training and evaluation but also introduce challenges related to class imbalance and noise.

4. **Critical Insights and Limitations**

Despite notable advances, limitations persist across the literature. These include difficulty in fine-grained geolocation, limited interpretability of deep learning models, bias caused by imbalanced humanitarian categories, and insufficient temporal–spatial visualization of evolving events. Moreover, many prior systems lack a fully integrated pipeline that combines classification, geolocation, credibility assessment, and event mapping.

(2) Integrating Machine Learning and Geospatial Data for Flood Hazard Mapping

Recent research demonstrates the growing role of machine learning and geospatial technologies in flood hazard assessment. By combining GIS, remote sensing, and ML techniques, recent studies have significantly improved the accuracy of flood susceptibility mapping compared to traditional models. However, several challenges remain, including limited model interpretability, insufficient sensitivity analysis, dependence on scarce field data, and difficulty capturing complex nonlinear relationships between flood-conditioning factors.

1. **Existing Methods and Techniques**

Earlier flood modeling studies relied on traditional statistical and machine learning methods such as Logistic Regression, Decision Trees, Frequency Ratio, and Weights of Evidence. While these approaches are computationally efficient, they struggle to represent nonlinear patterns, complex interactions, and large spatial datasets, resulting in limited generalization across different geographic regions. More recent research shows that advanced machine learning models—particularly ensemble methods such as Random Forest, Gradient Boosting, and XGBoost—consistently outperform traditional models by achieving higher predictive accuracy and robustness.

2. **Hybrid and Advanced Approaches**

Recent literature emphasizes hybrid and ensemble learning frameworks that integrate multiple ML algorithms with geospatial data. These approaches often combine remote sensing inputs, optimization-based feature selection, and ensemble classifiers to enhance model stability and performance. In addition, explainability techniques such as Boruta and SHAP are increasingly used to reduce the "black-box" nature of advanced models and to identify the most influential flood-conditioning factors. Compared to single-model approaches, hybrid frameworks provide better accuracy, improved reliability, and stronger generalization ability.

3. **Datasets and Evaluation**

Most flood hazard studies rely on multi-source geospatial datasets, including Sentinel-1 SAR imagery, digital elevation models, rainfall data, global flood inventories, and land-use and soil maps. These datasets are commonly processed using cloud platforms such as Google Earth Engine, with model performance evaluated through metrics such as accuracy and AUC. While large datasets improve model training and validation, they also introduce challenges related to data quality and spatial variability.

4. **Critical Insights and Limitations**

Despite substantial progress, key limitations remain in flood hazard modeling research. These include low interpretability of complex ML models, limited sensitivity analysis, reliance on specific topographic and rainfall variables, and weak transferability across regions. Moreover, many studies lack a unified and fully interpretable modeling pipeline that integrates feature selection, multiple ML techniques, and explainability.

(3) An Integrated Approach of Machine Learning, Remote Sensing and GIS Data for the Landslide Susceptibility Mapping.

1. **Primary Goal:** To assess landslide susceptibility as a single-hazard model in the Abbottabad District of Pakistan.
2. **Models and Tools:** Linear Regression (LiR), Logistic Regression (LoR), and Support Vector Machine (SVM).
3. **Data and Inputs:** The models relied on fourteen (14) distinct geological and environmental features such as slope characteristics, subsurface lithology, and the Normalized Difference Water Index as the primary inputs.
   The dataset was systematically partitioned into 70% for development and 30% for testing.
4. **Results:** Model effectiveness was quantified using the ROC curve and the Area Under the Curve (AUC) metric.
   The LiR model demonstrated superior performance, recording the highest AUC score of 0.88, which surpassed the performance of both the SVM and LoR models.
   The analysis confirmed that terrain gradient and lithological composition were the most influential predictors.

5. Critical Insights

5.1 Methodological Limitations:

- The classic models utilized ((LiR and SVM)) lack the capability for automatic extraction of complex features from spatial data.
- This represents a fundamental constraint, as the models heavily rely on the manual selection of the 14 previously mentioned factors.
- Scalability Issues: This lack of automated feature extraction significantly limits the model's ability to generalize and be applied effectively across different geographical regions.
- Researcher Recommendation (Gap): The researchers explicitly recommended that future work should focus on integrating Deep Learning (DL) techniques, specifically Convolutional Neural Networks (CNNs), to enhance feature extraction and boost overall results.
- Proposed Solution (Project Aim): Our project aims to develop an advanced ML module based on CNNs to surpass the established 0.88 AUC benchmark and provide a robust solution for disaster management authorities in aid distribution.

(4) A machine learning framework for multi-hazards modeling and mapping in a mountainous area.

1. **Primary Goal:** To address the challenge of multi-hazard modeling simultaneously.
2. **Hazards Covered:** The study encompassed the modeling of five specific hazards: landslides, floods, wildfire, snow avalanches, and land subsidence.
3. **Models and Tools:** Machine Learning models: Support Vector Machine (SVM), Generalized Linear Model (GLM), and Functional Discriminant Analysis (FDA).
4. **Data and Inputs:** These models relied on a comprehensive analysis of 39 distinct environmental and geographical factors.
5. **Results:**
    - Performance was assessed using the AUC metric.
    - This framework demonstrated acceptable overall accuracy, as the AUC value exceeded the 0.8 threshold for all five included hazards.
    - This success confirmed the feasibility of using Machine Learning models to generate unified multi-hazard maps and provide preliminary predictive data.
6. **Critical Insights**

    6.1 Methodological Limitations:
    - Despite achieving acceptable accuracy, the traditional models utilized suffer from a deficiency: they fail to model complex interactions and interdependencies between multiple hazards.
    - The models treat each hazard as an independent event, unable to capture the chain of interactions where one disaster exacerbates another (e.g., the impact of a snow avalanche on floods or landslides).

    6.2 Scalability/Operational Value Issues:
    - This failure to understand the inherent dependency between hazards significantly reduces the operational value of the resulting maps for advanced planning and crisis management.

    6.3 Proposed Solution (Project Aim):
    - It is essential to develop models that overcome these methodological limitations; hence, the focus must shift to Deep Learning (DL) techniques, specifically Convolutional Neural Networks (CNNs), to enable the model to comprehensively simulate these complex interactions and offer reliable, holistic solutions for disaster management.

(5) Flood Hazard and Risk Mapping by Applying an Explainable ML Framework Using Satellite Imagery and GIS Data

1. **The research aims** to create maps by fusing satellite imagery with (GIS) data.

2. **Datasets Used:**
   - **Satellite Imagery:** Utilizes Sentinel-1 radar imagery to identify areas.
   - **GIS-Based Data:** Geographically information integrates location components with attribute data, to show geographic context through maps.

3. **Techniques:**
   - **ML Models:** SVMs, NB, Random Forest, and Neural Networks.
   - **Risk Assessment:** rule-based approach based on District Authority's (AAW) from Flood Risk Management Plan to estimate Vulnerability and Exposure.

4. **Results:**
   - The RF achieved the best performance, with Precision and Recall scores 0.99, indicating a very high predictive accuracy on their dataset.
   - The feature importance analysis revealed that the most influential factors in determining flood hazard was Water Velocity and Water Depth were.

5. **Critical Insights**

5.1 Strengths

- The approach of **fusing satellite data with GIS** provides where the flood is while GIS provides how flood is behaved based on the terrain.
- The selection of **RF** was well-justified. Its resistance to overfitting, and its built-in ability to calculate feature importance.
- Utilities of **Satellite Image** have the ability to penetrate clouds and operate at night.

5.2 Limitations, Biases, and weaknesses:

- Accuracy is heavily dependent on the availability of a 0.5m resolution. This type is expensive, rare, and not available for most of the world.
- **Problem in Data Annotation Bias:** To train their supervised models, the authors needed an annotated dataset.
- The paper notes the satellite has a revisit time of 6 days too slow for monitoring the evolution of a rapid-onset disaster, in true real-time.

(6) Assessing and mapping multi-hazard risk susceptibility using a ML technique

1. **Primary Goal:** Multi-Hazard mapping system for floods, landslides, and forest fires.
2. **Datasets:**
   - 2.1. **Hazard locations:** Locations of 365 floods, 358 forest fires, and 179 landslides were obtained through intensive fieldwork using GP.
   - 2.2. **Non-hazard locations:** equal in number to the hazard events were randomly sampled for the modeling and validation process.
3. **Tools Used:**
   - 3.1. The Boruta algorithm (Feature Selection) was used to determine the importance of effective factors on the occurrence of floods, forest fires, and landslides.
   - 3.2. The Random Forest model was used to prepare the susceptibility maps for hazards.
4. **Results:**

   The three maps were combined to create the **Multi-Hazard Mapping. Boruta algorithm** identified the most important factors for each hazard. For instance, land use was most critical for floods, distance to residential areas for forest fires, and slope for landslides. The **Random Forest** model achieved high accuracy and AUC values of 0.834 for floods, 0.943 for forest fires, and 0.939 for landslides. The **multi-hazard map** revealed that 42.83% of the province is not susceptible to any hazard, but 2.67% is at risk from all three hazards simultaneously.

5. Critical Insights

5.1 Strength:

- The study combines the risks from floods, fires, and landslides, to create a **multi-hazard framework that** is highly valuable.

- The study covers a very **large-scale, regional** and diverse area 133.400.00 km², demonstrating that this ML approach is scalable.

- All three hazards, with the forest fire and landslide maps showing excellent AUC accuracy 0.943 and 0.939 which indicate **high predictive accuracy**.

5.2 Limitations, Biases, and issues:

- Compared to the landslide and forest fire maps, the **flood susceptibility map** has **more uncertainty**. The difficulty due to the challenge of locating flood events.

- Possibility of **biases** due to floods and landslides close to roads and cities are more likely to be reported than those in remote, difficult-to-reach places.

(7) Improving Social Media Geolocation for Disaster Response by Using Text From Images and ChatGPT

1. **Primary Goal:**

   The study aims to improve the geolocation of disaster events at the city and neighborhood levels by fusing Large Language Models (LLMs) with Optical Character Recognition (OCR).

2. **Datasets Used:**
   - Real-world Dataset: A dataset comprising four major disaster events, including the Haiti earthquake and European floods.
   - Social Media Data: Posts containing both text and images.

3. **Tools Used:**
   - Amazon Rekognition (OCR): Used to extract text embedded inside images (e.g., street signs, shop names).
   - ChatGPT (GPT-3.5) & GeoChatGPT: LLMs used to infer precise locations via question-answering prompt engineering.

4. **Techniques:**
   - Multimodal Fusion: The visual text extracted by OCR is concatenated with the user's post text and fed into the LLM.
   - Prompt Engineering: Utilizing a QA strategy to extract granular location details.

5. **Results:**

The combination of OCR-extracted text with post content significantly improved both the quantity and precision of detected locations.

6. **Comparison of Different Techniques**

   6.1 Overview:

   The study conducted a comparative ablation study testing different input configurations:

   6.2 **Technique: GeoChatGPT + OCR**

   - Configuration: Text + Visual Text
   - Performance Note: Highest Performance: Successfully identified granular locations (hospitals, streets) that were completely missed when relying on post text alone.

6.3 Technique: Standard ChatGPT

- Configuration: Text + Visual Text
- Performance Note: Improved over text-only but generally less specialized than GeoChatGPT.

6.4 Technique: Text-only Methods

- Configuration: Text Metadata only
- Performance Note: Often insufficient for precise geolocation.

7. Critical Insights

7.1 Strengths:

- Granularity: The approach successfully identifies specific landmarks like hospitals or streets which are missed when relying on post text alone.

7.2 Limitations, Biases, and Gaps:

- **Dependence on Visual Text (Major Limitation):** The method's success is strictly conditional on the presence of readable text within the image. It renders the improvement null for vast rural areas or natural landscapes without signboards.

- **Ambiguity and Hallucination (Bias):** LLMs are probabilistic and can hallucinate or misinterpret common names. The system requires a "human-in-the-loop" disambiguation mechanism to confirm correct sites.

- **Energy and Latency (Operational Gap):** Reliance on heavy LLMs implies higher energy consumption and potential latency compared to lightweight classifiers, posing a challenge for sustainable real-time systems

(8) VGI and Satellite Imagery Integration for Crisis Mapping of Flood Events

1. **Primary Goal:** The research aims to validate satellite-derived flood maps by creating a semi-automated framework that integrates Volunteered Geographic Information (VGI) with optical satellite imagery.

2. **Datasets Used:**
- VGI Data: Geolocated social media photographs from Flickr, Twitter, and YouTube.
- Satellite Imagery: Optical imagery from Landsat-7 and Sentinel-2.
- Reference Data: Official radar-based flood maps (e.g., Copernicus EMS, HASARD).

3. **Tools Used:**
- Spectral Angle Mapper (SAM): A classifier used to detect floodwater in satellite imagery.
- QField: A mobile application prototype proposed for structured field data collection.

4. **Techniques:**
- ROI Reconstruction: Calculating camera parameters (view angle, orientation) from photos to project them onto the map.
- Ground Truth Generation: Using VGI samples to train the classifier instead of relying solely on experts.

5. **Results:**
- The workflow achieved an Overall Accuracy (OA) ranging between 87% and 93% for successful test cases (UK and US floods).

6. Comparison of Different Techniques

The study compared the proposed VGI-based classification against official standards:

6.1 Technique: VGI + SAM Classifier

- Context: Proposed Method
- Performance Metric: Overall 87% - 93% Accuracy (in successful scenarios).

6.2 Technique: Official Radar Maps

- Context: Baseline/Reference
- Performance Metric: Used as the ground truth for cross-referencing.

6.3 Technique: Traditional Surveys

- Context: Conventional Method
- Performance Metric: The study proved social media imagery can effectively substitute these expensive surveys.

7. Critical Insights

7.1 Strengths:

- Cost-Efficiency: Demonstrated that properly processed social media imagery can substitute traditional, expensive ground surveys for training satellite classifiers**.**

7.2 Limitations, Biases, and Gaps:

- The "Cloud Cover" Limitation (Major Limitation): Optical satellites cannot see through clouds. This rendered the system ineffective during the peak of the storm in the US Wilmington case
- Temporal Mismatch (Operational Gap): Social media is real-time, while satellites have fixed revisit times. If the satellite pass does not coincide with the peak flooding in photos, the map becomes inaccurate.
- Data Scarcity and Pre-processing (Weakness): Extracting usable VGI is labor-intensive; out of hundreds of posts, only a handful (e.g., 13 in the UK case) were accurate enough. The heavy reliance on manual filtering contradicts the need for "rapid" response.

(9) Detecting Emergency Rescue Requests on Twitter Detecting Emergency Rescue Requests on Twitter

1. **Primary Goal:**
   Identify real emergency rescue requests on Twitter during hurricanes.
2. **Models and Tools:**
   - Rule-based filtering (keyword lists, address patterns).
   - Machine-learning classifiers using BERT embeddings for contextual understanding.
3. **Data and Inputs:**
   - Tweets collected using hurricane-related hashtags, location keywords, and U.S. city names.
   - Manually labeled tweets based on two criteria: explicit help request + location mention.
4. **Results:**
   - The hybrid approach (rules + BERT) achieved higher accuracy than rule-based methods alone.
   - Successfully detected linguistic cues such as “trapped,” “urgent,” or specific addresses.
5. **Limitations:**
   - Restricted to English tweets only.
   - Requires extensive manual labeling, which is time-consuming.
6. **Biases:**
   - Twitter usage varies among demographic groups, causing uneven model performance.
   - Informal writing styles may cause the model to miss genuine rescue tweets.
7. **Scalability Issues:**
   - High tweet volume during disasters increases processing load.
   - Real-time classification requires strong computational resources.
8. **Gaps:**
   - Lack of interpretability in BERT-based decisions.
   - No handling of multilingual rescue messages.
   - Limited testing across different disaster types.

(10) DAHiTrA for Building Damage Assessment

1. **Primary Goal:**
   Automatically estimate building damage using pre- and post-disaster satellite imagery.

2. **Models and Tools:**
   - UNet encoder for multi-resolution feature extraction.
   - Transformer encoders/decoders for global attention and feature alignment.
   - Multi-scale up-sampling module to generate high-resolution damage maps.

3. **Data and Inputs:**
   - xBD dataset (large-scale global disaster imagery).
   - Ida-BD dataset (high-resolution images after Hurricane Ida).
   - Paired before/after satellite images.

4. **Results:**
   - Outperformed baseline models (Siamese UNet, RescueNet, BDANet, BiT).
   - Achieved strong performance in building segmentation and damage-level classification.
   - Successfully adapted to a new dataset (Ida-BD) with minimal fine-tuning.

5. **Limitations:**
   - Sensitive to variation in lighting, angle, and image quality.
   - Requires precise alignment between pre- and post-disaster images.

6. **Biases:**
   - Datasets overrepresent certain regions and disaster types.
   - May not generalize well to low-resource or non-Western environments.

z

7. **Scalability Issues:**
   - Transformer layers require significant memory and computational power.
   - Processing large high-resolution satellite images can be slow.

8. **Gaps:**
   - Limited interpretability of the model's internal decision-making.
   - No integration with real-time disaster monitoring systems.
   - Does not account for non-building infrastructure damage.

Primary Goal	To build an integrated ML pipeline that automatically classifies disaster-related tweets, geolocates them, filters unreliable posts, and produces credible spatio-temporal maps of disaster events.	To develop a comprehensive ML-based framework for high-accuracy flood hazard/susceptibility mapping by integrating multi-source geospatial data, advanced ML models, feature selection, and explainability techniques.
Models and Tools	BERT-based tweet classifier; Named Entity Recognition (NER) for place names; Google Maps Geocoding API for coordinates; graph-based clustering for event detection and noise filtering; baselines include CNN, LSTM, GRU, Logistic Regression, and SVM with TF-IDF.	Advanced ML algorithms such as Random Forest, XGBoost, Gradient Boosting, AdaBoost, LightGBM, CatBoost, DeepBoost, Artificial Neural Networks, and neuro-fuzzy models; ensemble strategies (bagging, RF, GBM, rFerns); feature selection via OLS regression, multicollinearity tests, and nature-inspired optimizers (PSO, GA, GSO, GWO, HHO); explainability tools like Boruta and SHAP; implemented within GIS/GEE environments.
Data and Inputs	Large-scale Twitter data collected during major disasters (e.g., hurricanes), filtered by disaster-related keywords; raw tweet text plus limited metadata; datasets cleaned by removing retweets, replies, and obviously irrelevant content.	Remote sensing and geospatial layers including Sentinel-1 SAR (flood extent), Sentinel-2 NDVI, ASTER DEM (elevation, slope, curvature), rainfall products (e.g., TerraClimate), soil, land cover, and geomorphology maps; Global Flood Database and GPS field surveys for flood inventory; around 15 flood-conditioning factors, typically split into 70% training and 30% testing.
Results	Fine-tuned BERT significantly outperforms traditional ML and earlier deep models in humanitarian tweet classification; NER + geocoding provides more precise locations than rule-based methods; graph clustering effectively removes isolated/noisy tweets and groups messages describing the same real-world events, enabling more reliable disaster maps.	Ensemble and tree-based models (especially RF and XGBoost) achieve very high predictive performance (AUC often ≥0.89 and up to ~0.99 in some case studies), with strong sensitivity and specificity; optimization-based feature selection improves model stability and accuracy; Boruta/SHAP analyses identify the most influential factors (e.g., elevation, rainfall, distance to rivers); resulting hazard maps provide fine spatial differentiation of flood-prone areas.

Limitations	Geolocation remains coarse when tweets contain vague or ambiguous place references; class imbalance (e.g., rare categories like missing persons) introduces bias; model decisions are not easily interpretable; credibility still depends on indirect signals rather than verified ground truth; strong dependence on Twitter as a single data source.	Many models still behave as "black boxes" despite post-hoc explainability; limited temporal coverage and quality of flood observations constrain validation; strong dependence on DEM and rainfall quality can introduce systematic bias; models often tuned for a specific basin, so generalization to other regions is uncertain
Scalability Issues	Running BERT on millions of tweets, repeated geocoding API calls, and graph clustering over large networks is computationally expensive; external geocoding APIs add latency, quota constraints, and cost; near–real-time deployment requires substantial optimization and infrastructure.	Training multiple advanced models and running optimization algorithms on high-resolution rasters requires significant computational power and storage; heavy reliance on cloud platforms (e.g., GEE) and large geospatial datasets may limit use in resource-constrained settings; frequent updates of hazard maps as new data arrive can be computationally and operationally costly.

Table 1 : Comparison table between the first and second research

Primary Goal	Aimed to evaluate the likelihood of landslides as a specific, single-hazard model within the Abbottabad District of Pakistan	Focused on solving the complexity of modeling multiple hazards simultaneously, covering five distinct disasters (landslides, floods)
Models and Tools	Utilized three standard Machine Learning algorithms: Linear Regression (LiR), Logistic Regression (LoR), and Support Vector Machine (SVM)	Constructed a framework using three traditional models: Support Vector Machine (SVM), Generalized Linear Model (GLM), and Functional Discriminant Analysis (FDA).
Data and Inputs	Relied on 14 specific geological and environmental indicators (e.g., slope, lithology) as key inputs, split into 70% for training and 30% for testing.	Based on a comprehensive analysis of 39 different environmental and geographical factors.
Results	The Linear Regression (LiR) model achieved the best performance, reaching a high AUC score of 0.88	The framework showed acceptable accuracy, with AUC values exceeding the 0.8 threshold for all five hazards analyzed
Limitations	The traditional models used lack the ability to automatically extract complex features from spatial data, relying heavily on manual selection of factors.	The models treat each hazard as an isolated event and fail to capture the complex chain reactions or interdependencies between different hazards.
Scalability Issues	The lack of automated feature extraction makes it difficult to generalize the model or apply it effectively to different geographic regions.	By missing the dependencies between hazards, the resulting maps offer reduced operational value for advanced crisis management and planning.

Table 2 : Comparison table between the first and second research

Primary Goal	The research aims to create maps by fusing satellite imagery with (GIS) data	Combine risks of floods, landslides, and forest fires into multi-hazard mapping.
Models and Tools	**ML Models used** SVMs, NB, Random Forest, and Neural Networks. Rule-based approach to estimate Vulnerability and Exposure.	The **Boruta algorithm was** used to determine the importance of effective factors. The **Random Forest** model was used to prepare the susceptibility maps for hazards.
Data and Inputs	Utilizes **Satellite Imagery** Sentinel-1 radar to identify areas. **Geographically information** integrates location components with attribute data, to show geographic context through maps.	**Hazard locations** of 365 floods, 358 forest fires, and 179 landslides were obtained using GPS and province reports. **Non-hazard locations** equal in number to the hazard events were randomly sampled for the modeling and validation process.
Results	The RF achieved the best performance, with Precision and Recall scores 0.99, indicating a very high predictive accuracy on their dataset. The feature importance analysis revealed that the most influential factors in determining flood hazard was Water Velocity and Water Depth were.	**After** identifying important factors for each hazard. Thus, land use was critical for floods, distance to residential areas for forest fires, and slope for landslides. The **RF** achieved high accuracy and AUC values of 0.83 for floods, 0.94 for forest fires, and 0.93 for landslides. The **map** revealed that 42.8% of the province isn't susceptible to any hazard, but 2.6% is at risk from all hazards.
Limitations & Scalability Issues	Accuracy is heavily dependent on 0.5m resolution that is expensive, rare and not available. The authors needed an annotated dataset which causes bias. Also satellite radar has a revisit time of 6 days which is too slow for monitoring the evolution of disaster.	Compared to landslide and forest fire maps, the **flood map** has **more uncertainty**. The difficulty due to challenges of locating flood events. **Possibility of biases** due to floods and landslides close to roads and cities likely to be reported than those which are difficult-to-reach places.

Table 3 : Comparison table between the first and second research

Primary Goal	The study aims to better locate disaster events in cities and neighborhoods using LLMs and OCR.	Validating satellite flood maps using VGI and optical imagery in a semi-automated framework.
Models and Tools	Amazon Rekognition (OCR): Used to extract text embedded inside images (e.g., street signs, shop names).While ChatGPT (GPT-3.5) & GeoChatGPT: LLMs used to infer precise locations via question-answering prompt engineering.	Spectral Angle Mapper (SAM): A classifier used to detect floodwater in satellite imagery. And QField: A mobile application prototype proposed for structured field data collection.
Data and Inputs	• Real-world Dataset: A dataset comprising four major disaster events, including the Haiti earthquake and European floods. • Social Media Data: Posts containing both text and images.	• VGI Data: Geolocated social media photographs from Flickr, Twitter, and YouTube. • Satellite Imagery: Optical imagery from Landsat-7 and Sentinel-2. • Reference Data: Official radar-based flood maps (e.g., Copernicus EMS, HASARD).
Results	The combination of OCR-extracted text with post content significantly improved both the quantity and precision of detected locations.	The workflow achieved an Overall Accuracy (OA) ranging between 87% and 93% for successful test cases (UK and US floods).
Limitations	• Works only when readable text exists in images. • LLMs can misinterpret locations and need human confirmation. • High energy use and delay limit real-time deployment.	Clouds block optical satellite visibility. Timing mismatch between social media and satellite passes. Limited usable VGI and heavy manual filtering.
Scalability Issues	High computational cost of OCR + LLM limits large-scale real-time deployment.	Scaling the system to large geographic areas is limited by satellite revisit times, cloud coverage, and the need for manual VGI filtering.

Table 4 : Comparison table between the first and second research

Primary Goal	Identify real emergency rescue requests on Twitter during hurricanes.	Automatically estimate building damage using pre- and post-disaster satellite imagery.
Models and Tools	• Rule-based filtering (keyword lists, address patterns). • Machine-learning classifiers using BERT embeddings for contextual understanding.	• UNet encoder for multi-resolution feature extraction. • Transformer encoders/decoders for global attention and feature alignment. • Multi-scale up-sampling module to generate high-resolution damage maps.
Data and Inputs	• Tweets collected using hurricane-related hashtags, location keywords, and U.S. city names. • Manually labeled tweets based on two criteria: explicit help request + location mention.	• xBD dataset (large-scale global disaster imagery). • Ida-BD dataset (high-resolution images after Hurricane Ida). • Paired before/after satellite images.
Results	• The hybrid approach (rules + BERT) achieved higher accuracy than rule-based methods alone. • Successfully detected linguistic cues such as “trapped,” “urgent,” or specific addresses.	• Outperformed baseline models (Siamese UNet, RescueNet, BDANet, BiT). • Achieved strong performance in building segmentation and damage-level classification. • Successfully adapted to a new dataset (Ida-BD) with minimal fine-tuning.
Limitations	• Restricted to English tweets only. • Requires extensive manual labeling, which is time-consuming.	• Sensitive to variation in lighting, angle, and image quality. • Requires precise alignment between pre- and post-disaster images.
Scalability Issues	• High tweet volume during disasters increases processing load. • Real-time classification requires strong computational resources.	• Transformer layers require significant memory and computational power. • Processing large high-resolution satellite images can be slow.

Table 5 : Comparison table between the first and second research

5. Gap Analysis and Justification

Gaps and Limitations Across the Reviewed Studies

Across all the studies, the same structural weaknesses keep showing up, even if the data sources differ. Most models fail to generalize beyond the narrow environments they were trained on, whether the issue is English only Twitter datasets, flood models restricted to a single basin, landslide studies tied to local geological inputs, or satellite pipelines that depend on rare 0.5-meter imagery. These systems tend to break the moment they face new regions, languages, hazards, or social-media behaviors, and this lack of transferability is reinforced by data scarcity, heavy annotation requirements, and demographic biases that cluster reports around cities while leaving remote areas underrepresented. Even when large datasets are available, they suffer from noise, misinformation, bot activity, informal writing styles, and uneven reporting patterns that distort the ground truth. Satellite-based approaches introduce their own blind spots: optical sensors fail under cloud cover, revisit cycles lag behind fast-moving disasters, and both pre/post-event alignment and lighting variation reduce reliability. Whether dealing with social media, satellite images, or environmental layers, all pipelines face temporal mismatch between human reporting and remote-sensing acquisition, producing maps that are often incomplete or outdated.

Methodologically, traditional ML models still rely on hand-crafted features and cannot capture complex spatial relationships or multi-hazard interactions, while deep models tend to operate as opaque black boxes with limited interpretability for decision-makers. Even advanced Twitter classifiers or transformer-based satellite models struggle with credibility scoring, uncertainty estimation, multilingual handling, and the integration of heterogeneous signals. Several studies highlight that their systems treat hazards independently, ignoring the way disasters trigger or amplify one another, and others emphasize that purely geospatial pipelines lack any connection to human observations. Additionally, heavy computational loads, high memory requirements, and the cost of high-resolution data make real-time or large-area scaling difficult. Across all domains, the absence of unified, end-to-end frameworks leads to fragmented outputs that reflect isolated data sources rather than the full operational reality of unfolding disasters.

Justification for the Proposed Project

These recurring weaknesses make the need for an integrated, interpretable, and multi-source ML system unavoidable. Because current models either overfit to narrow datasets or operate as black boxes without meaningful validation, our project addresses the problem by merging geospatial flood indicators with real-time social-media observations, allowing each source to confirm, refine, or challenge the other. This fusion directly overcomes the credibility gaps, temporal blind spots, and spatial inconsistencies highlighted across the studies. Using deep learning for automatic feature extraction resolves the limitations of classical ML models that depend on manually selected factors, while explainable ML techniques—such as SHAP, transparent attribution layers, and uncertainty scoring—resolve the interpretability and trust issues that emergency teams consistently struggle with. Integrating multi-hazard reasoning avoids the common pitfall of treating disasters in isolation, and incorporating multilingual, modular components improves adaptability across regions and data ecosystems. By designing the system for operational use rather than academic benchmarking, the framework shifts from static susceptibility maps and delayed satellite snapshots to dynamic, time-aware intelligence grounded in both physical models and human-generated reports. In short, the proposed project is not just an improvement, it is a direct, evidence-driven answer to the core gaps, biases, and constraints repeatedly documented across all the reviewed studies.

6. ML Module Proposal

6.1 Objectives

The proposed Machine Learning module aims to automatically detect and map disaster-affected areas by comparing pre-disaster and post-disaster satellite imagery. Its primary objective is to generate an accurate, automated damage map that highlights critical zones requiring urgent intervention. This module seeks to improve the speed, precision, and scalability of disaster assessment, enabling emergency response teams to make informed decisions based on reliable geo-spatial intelligence.

6.2 Approach

The system will use a Random Forest Classifier as a supervised machine learning technique to identify damaged areas. This algorithm was chosen because it:

- Handles a large number of features extracted from satellite images
- Captures complex changes between pre- and post-disaster imagery
- Provides stable and interpretable results
- Maintains high accuracy even when the data is noisy or partially incomplete

Additionally, the top-performing research study we reviewed also used this algorithm, and it achieved the highest accuracy compared to other methods, which supports our decision to rely on it in this project.

The model will be trained using labeled examples of damaged and undamaged regions, using features derived from spectral values, change-detection metrics, and texture descriptors.

6.3 Workflow Diagram:

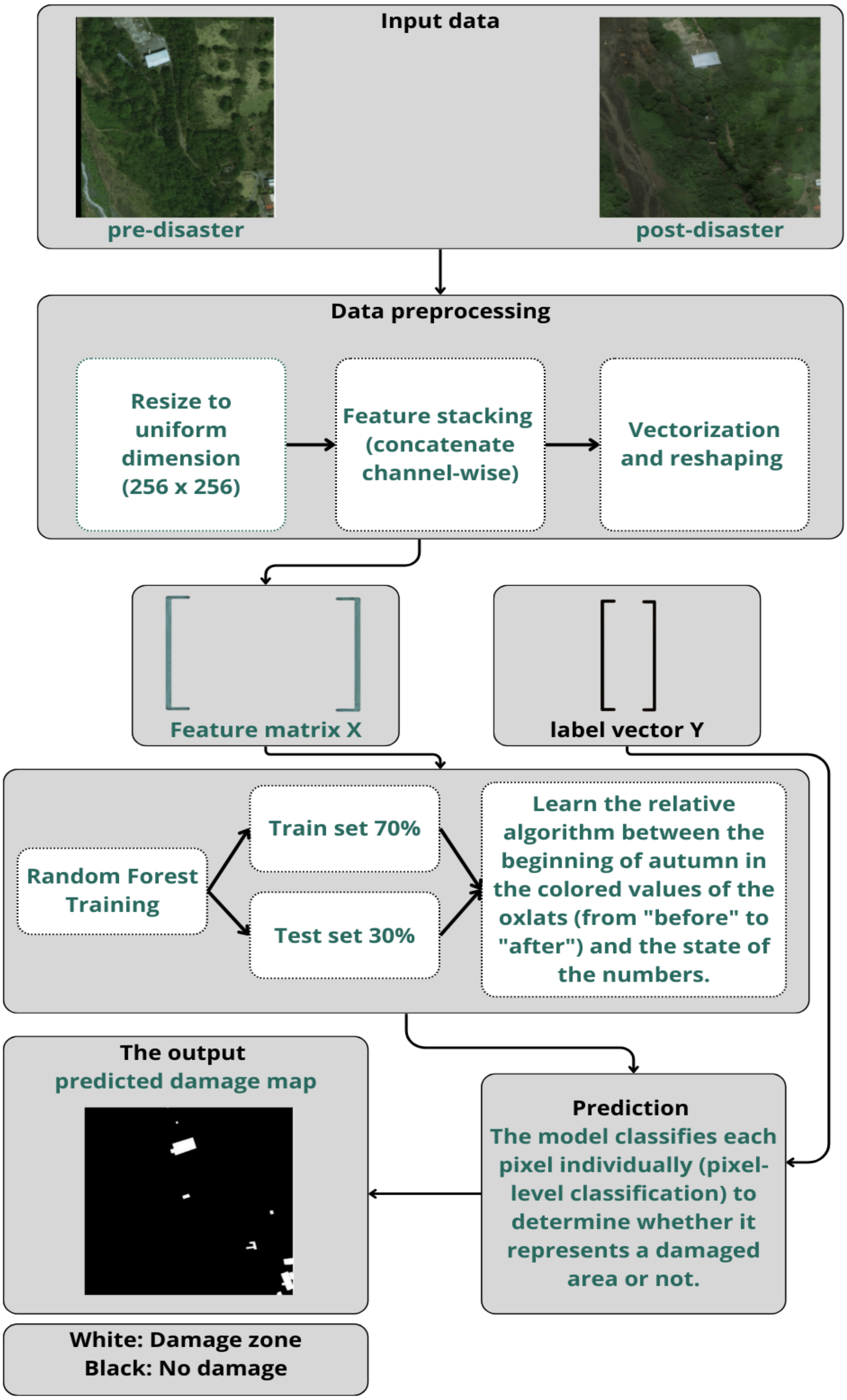

Figure 1. flow chart of multi hazard mapping

6.4 Expected Benefits

By detecting and mapping affected areas from satellite imagery, the system used random forest to create a damage map in minutes rather than days by analyzing large datasets. The system can process images over a wide area without increasing processing time. However human analysts may become exhausted by the huge amount of satellite imagery. Unlike manual analysis, which can be subjective and differ between analysts, the model applies the same criteria across all images.

This reduces human error and provides a trustworthy basis for decision-making that leads to standardized and reliable systems.

Emergency response teams can obtain situational awareness much more quickly due to this speed improvement, which increases relief operations.

The result is a single, clear, and useful damage map rather than just the collection of pictures. The system transforms complex visual data into a simple, understandable tool that emergency planners can use immediately by indicating critical zones in red "Damage Zone" and unaffected areas in black "No Damage".

It enables emergency response teams to make well-informed, data-driven decisions in almost real-time by offering a quick and trustworthy summary of the disaster's effects.

UMM AL-QURA UNIVERSITY

7.Conclusion

In disaster management studies, accurate risk and hazards maps are crucial for preparedness and mitigating extreme disaster situations. In this paper we successfully designed and justified a machine learning module by proposing a framework based on a Random Forest classifier. The core of our methodology involves comparing pre- and post-disaster satellite imagery, with the Random Forest model trained to learn the complex patterns and feature changes such as changes in structural, spectral indices, and texture that indicate damage. This process is designed to transform raw satellite imagery into essential geographic information and produce an actionable damage map much faster than human analysts. We have demonstrated a robust and effective approach to detecting and mapping disaster affected areas using new data compared to our review of the literature.

This field must create interpretable, and multi source ML frameworks. Future research must focus on data fusion merging disparate sources like satellite flood indicators and real time social media observations to resolve temporal blind spots. The use of Deep Learning for feature extraction is necessary to overcome the limitations of manually selected factors in classical models. In conclusion, research needs to focus on fundamental gaps, biases, and limitations that have been repeatedly noted throughout the reviewed studies by giving priority to dynamic, multi-hazard reasoning.

8. References

[1] C. Singha, V. K. Rana, Q. B. Pham, D. C. Nguyen, and E. Łupikasza, "Integrating machine learning and geospatial data analysis for comprehensive flood hazard assessment," Environmental Science and Pollution Research, vol. 31, no. 35, pp. 48497–48522, Jul. 2024, doi: 10.1007/s11356-024-34286-7.

[2] C. Fan, F. Wu, and A. Mostafavi, "A hybrid machine learning pipeline for automated mapping of events and locations from social media in disasters," IEEE Access, vol. 8, pp. 10478–10490, 2020, doi: 10.1109/ACCESS.2020.2965550.

[3] I. Ullah et al., "An integrated approach of machine learning, remote sensing, and GIS data for the landslide susceptibility mapping," Land, vol. 11, no. 8, Art. no. 1265, Aug. 2022, doi: 10.3390/land11081265.

[4] S. Yousefi, H. R. Pourghasemi, S. N. Emami, S. Pouyan, S. Eskandari, and J. P. Tiefenbacher, "A machine learning framework for multi-hazards modeling and mapping in a mountainous area," Scientific Reports, vol. 10, no. 1, Art. no. 12144, Jul. 2020, doi: 10.1038/s41598-020-69233-2.

[5] G. Antzoulatos et al., "Flood hazard and risk mapping by applying an explainable machine learning framework using satellite imagery and GIS data," Sustainability, vol. 14, no. 6, Art. no. 3251, Mar. 2022, doi: 10.3390/su14063251.

[6] H. R. Pourghasemi et al., "Assessing and mapping multi-hazard risk susceptibility using a machine learning technique," Scientific Reports, vol. 10, no. 1, Art. no. 3203, Feb. 2020, doi: 10.1038/s41598-020-60191-3.

[7] H. B. Firmansyah, V. Lorini, M. O. Mulayim, J. Gomes, and J. L. Fernandez-Marquez, "Improving social media geolocation for disaster response by using text from images and ChatGPT," in Proc. 11th Multidisciplinary Int. Social Networks Conf. (MISNC), Bali, Indonesia, Aug. 2024, pp. 67–72, doi: 10.1145/3675669.3675696.

[8] A. Vavassori, D. Carrion, B. Zaragozi, and F. Migliaccio, "VGI and satellite imagery integration for crisis mapping of flood events," ISPRS International Journal of Geo-Information, vol. 11, no. 12, Art. no. 611, Dec. 2022, doi: 10.3390/ijgi11120611.

[9] W. Khallouli, J. Li, J. Huang, G. Rabadi, and S. Kovacic, “An integrated machine learning approach for identifying emergency rescue messages on social media during natural disasters,” Social Network Analysis and Mining, vol. 15, no. 1, Art. no. 50, 2025, doi: 10.1007/s13278-025-01462-7.

[10] N. Kaur, C. C. Lee, A. Mostafavi, and A. Mahdavi-Amiri, “Large-scale building damage assessment using a novel hierarchical transformer architecture on satellite images,” Computer-Aided Civil and Infrastructure Engineering, vol. 38, no. 15, pp. 2072–2091, 2023, doi: 10.1111/mice.13083.

[11] Stats Wire, “Random Forest Classification | Machine Learning | Python,” YouTube, 2021. [Online Video]. Available: https://youtu.be/3NdH3egUjpM (https://www.google.com/search?q=https://youtu.be/3NdH3egUjpM). (https://www.google.com/search?q=https://youtu.be/3NdH3egUjpM) [Accessed: Dec. 11, 2025].

[12] Kunaal Naik | AI Agents, “Random Forest Hyperparameter Tuning using RandomisedSearchCv | Machine Learning Tutorial,” YouTube, 2020. [Online Video]. Available: https://youtu.be/SctFnD_puQI (https://www.google.com/search?q=https://youtu.be/SctFnD_puQI). (https://www.google.com/search?q=https://youtu.be/SctFnD_puQI) [Accessed: Dec. 11, 2025].

[13] “Lecture 14: Random Forest,” Machine Learning Course Material, Dept. of Computer Science, Umm Al-Qura University, Makkah, Saudi Arabia, 2025. [Online]. Available: https://lms.uqu.edu.sa (https://lms.uqu.edu.sa/). (https://lms.uqu.edu.sa/)

[14] OpenAI, “ChatGPT (GPT-4 model),” [Large language model], 2025. [Online]. Available: https://chat.openai.com (https://chat.openai.com/). (https://chat.openai.com/) [Accessed: Dec. 11, 2025].

[15] Google, “Gemini (Advanced model),” [Large language model], 2025. [Online]. Available: https://gemini.google.com (https://gemini.google.com/). (https://gemini.google.com/) [Accessed: Dec. 11, 2025].