

Ideation Phase

Define the Problem Statements

Date	18 october 2022
Team ID	PNT2022TMID43250
Project Name	Natural Disasters Intensity Analysis and Classification using Artificial Intelligence
Maximum Marks	2 Marks

Problem Statements

During the collection and handling of data, it is important to consider:

- (a) biases in training/testing datasets,
- (b) new distributed AI technologies within the data domain and
- (c) ethical issues.

In terms of biases in training/testing datasets, it is important to ensure that data are correctly sampled and that there is sufficient representation of each pattern for the problem in question.

Consider, for instance, the challenge of building a representative dataset containing examples of extreme events (which are, by nature, rare).

Also, imagine the possible costs of failing to provide appropriate data, for instance, wrong predictions or biased outcomes.

Once we have ensured that a dataset is not biased, we also need to decide how to integrate new distributed AI technologies within the data domain. Strategic modifications on the construction of space-based instruments like multiple small satellites³ and the introduction of edge computing (Nikos et al., 2018) have resulted in petabytes of data. Because AI relies on data transmission and computation of complex machine learning algorithms, centralized data processing and management can impose difficulties. On the one hand, real-time disaster applications require strong partnerships and data sharing between countries (recall the tsunami use case; Figure 4). On the other hand, ML algorithms are often operated in a centralized fashion, requiring training data to be fused in data servers. A centralized approach can also introduce additional challenges, such as privacy risks to personal and country-specific data. Furthermore, centralized data processing and management can limit transparency, which could lead to a lack of trust from end-users as well as difficulty in complying with regulations (e.g., GDPR).

Another data-related challenge is tied to ethical considerations. These centre on how AI-driven tools ought to be implemented from development to deployment, ensuring, for example, that socio-economic biases in underlying data are not propagated through the models developed by the system. Such principles are championed so that potential harms associated with AI, such as underrepresentation due to bias (either technical or human-based),

can be mitigated – if not removed, and so that the benefits of AI can be realized for all, especially those made more vulnerable by the impacts of natural hazards.⁴

After a dataset has been curated, we also need to consider challenges at the model development stage. Here, we focus on the computational demands and transparency. AI models tend to rely on complex structures and, as a result, can be computationally expensive to train. For example, the VGG16 model (Simonyan and Zisserman, 2015), which is used for image classification, has approximately 138 million trainable parameters. Training models of this size requires large and expensive computing capacity, which is not always accessible.

Once an AI model is developed, it is important that the results are humanly understandable and acceptable. This can be challenging to obtain because there is no general out-of-the-box human-machine interface that provides information about how and why certain decisions are made by the AI model.

Consequently, many researchers are working toward developing trustworthy AI solutions. In modelling and model evaluation, for instance, it is important to have a precise formulation of the problem and the requirements and expectations of the AI-based solution. Only then can a suitable model and learning strategy be developed to tackle the problem. Moreover, understanding the precise setup also helps in choosing and developing corresponding evaluation criteria.

For an AI-based model that is deemed ready for operational implementation, it is important to consider the aforementioned – data and model development-related – challenges as well as user notification challenges.

These are explored using AI-based communications technologies. To improve and facilitate interpretation of AI model outputs, these need to be translated and visualized according to end-user needs.

Therefore, it is critical that stakeholders – from local communities to emergency system managers – and NGO disaster response leaders be included in the design and evaluation of alert and early warning systems, forecasts, hazards maps, decision support systems, dashboards, chatbots and other AI-enhanced communications tools.

Timely feedback and evaluation of AI model insights from disaster responders is essential to improve the quality and precision of insights.

Transparency into the data sources ingested, the frequency of data refresh and the algorithms used for the communication tools is essential to develop trust and refinement of machine learning-based recommendations.

As with traditional modelling approaches, conveying confidence levels, uncertainties and limitations of an AI-enhanced system in an understandable way is crucial for informed decision-making.

Ultimately, trust in timely and fully transparent AI-based communications tools is the biggest challenge to be overcome.

This requires effective collaboration among experienced disaster responders, AI developers, geoscientists, regulators, government agencies, NGOs, telecom companies and others, to meet the needs of all stakeholders. Each disaster type is unique, and each region has different vulnerabilities and resiliency levels.