

# Accelerating Crisis Response: Automated Image Classification for Geolocating Social Media Content

Hafiz Budi Firmansyah

*Centre Universitaire d'Informatique  
University of Geneva  
Geneva, Switzerland  
hafiz.firmansyah@etu.unige.ch*

Jose Luis Fernandez-Marquez

*Centre Universitaire d'Informatique  
University of Geneva  
Geneva, Switzerland  
josemanuel.fernandez@unige.ch*

Mehmet Oguz Mulayim

*Artificial Intelligence Research Institute, IIIA  
Spanish National Research Council, CSIC  
Cerdanyola del Vallès, Spain*

Jorge Gomes

*Virtual Operations Support Teams  
VOST Europe  
Santarém, Portugal*

Valerio Lorini

*Joint Research Center  
European Commission  
Ispra, Italy*

**Abstract**—In the immediate aftermath of natural or man-made disasters, social media plays an essential role in assessing the impact of the event. The images from social media demonstrated the potential to accelerate the response to a crisis. However, finding the exact location of relevant social media images remains a problem for both humans and computer systems.

This study presents an automated image classifier aimed at accelerating crowdsourced geolocation. The classifier is trained with data annotated by crisis risk experts and predicts the difficulty in geolocating a photo. The experimental results demonstrate that the proposed approach can predict the geolocating difficulty, thus potentially speed up the geolocation process by presenting volunteers images that are easy to geolocate.

**Index Terms**—Social media, geolocating, disaster response

## I. INTRODUCTION

Social media platforms have the potential to enhance crisis response times. However, employing social media data for damage assessment comes with non-trivial challenges. For instance, many social media platforms remove the Global Position System (GPS) meta-data (i.e. the location) attached to the photo due to privacy reasons. Consequently, only up to 3% of images are geolocated in Twitter [1]. Identifying the location of relevant photos with evidence of damage to organize the first response efficiently is still challenging. While several techniques have been proposed to automatically geolocate social media content, including images [2] and text [3], their predictions need to be more precise to enable their implementation by practitioners in the field. A helpful contribution to alleviate this issue is to develop an automatic geolocation solution that serves as an initial estimation. The

primary objective of extracting location from social media images is to generate a disaster map, aiding the first responders to be better informed about the impact and coordinate the response accordingly.

Through the combination of Crowdsourcing and Artificial Intelligence approaches, this study presents an automated image classifier aimed at enhancing crowdsourced geolocation. The presented approach is a preliminary step towards decreasing the overall time of geolocating social media images originating from regions impacted by a crisis. The study leverages previous research activities involving social media images for the disaster risk management domain to serve emergency response as a framework [4], platform [5], [6], and disaster map [7].

This paper is structured as follows: Section II discusses related work. In Section III, we present the dataset, methodology, and experiment. Section IV reports the experiment results with social media data from four disaster events. Finally, Section V concludes the study and provides future work direction.

## II. RELATED WORK

Geolocation of social media data has been addressed in the literature by several studies, which use images, text or hybrid approaches. In an image-based approach, Murgese et al. [2] suggested a solution for estimating images' locations by implementing Focal Modulation Networks to predict the geolocation. The authors used images from Flickr<sup>1</sup> and Mapillary<sup>2</sup> to train and evaluate their image-based approach. Around 70% of Mapillary data, which had street-level imagery similar to Google Street View, was successfully pinpointed at the neighborhood level. However, the suggested approach did not involve social media data, and the precision of geolocation prediction was confined to the neighborhood level.

Some works attempt to use text-based geolocation. For example, Scalia et al. [3] proposed a context-aware approach

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ASONAM '23, November 6-9, 2023, Kusadasi, Turkey

© 2023 Association for Computing Machinery.

ACM ISBN 979-8-4007-0409-3/23/11...\$15.00

<http://dx.doi.org/10.1145/3625007.3627831>

<sup>1</sup><https://www.flickr.com/>

<sup>2</sup><https://www.mapillary.com>

to geolocate emergency-related social media posts. The system applied post meta-data and post network of relationships. The experimental results demonstrated that half of the images were predicted within 12 km from the actual location. The research which is focused on text-based geolocation prediction commonly incorporates textual content derived from social media posts. Another option is to extract text from an image and utilize this extracted text to retrieve the location. For instance, Firmansyah et al. [8] suggested a text-extraction pipeline to automatically infer the geolocation of social media images. In their study, the text within a social media image served as input for a text-based geolocation prediction algorithm. The experimental results revealed that the percentage of original images successfully geolocated within the bounding boxes (the impacted area) ranged from 9% to 11% of the initial dataset, depending on the geolocation algorithm used. Even though the work presents a great potential to be combined with other approaches, it is still in the preliminary phases of development and it currently presents a country-level precision.

In a hybrid technique, Ravi-Shankar et al. [9] presented a platform called Crowd4EMS. The presented approach analyzes and geolocates social media information by combining crowdsourcing and automatic methods. They leveraged data related to the Amatrice Earthquake in 2016, coming from Flickr, Twitter, and YouTube to evaluate the platform. The work showed that the suggested approach could better support the crowdsourcing communities in providing high-precision geolocation in the context of disaster response. However, the work also shows that crowdsourcing the geolocation of social media images is a complex and time-consuming task, which produces unpredictable delays that could prevent the information from being available in the immediate aftermath of the event.

This paper suggests that the timing and accuracy of the overall geolocation process of social media posts using crowdsourcing and automatic methods can be improved by prioritizing those images deemed easy to geolocate. This process involves the use of an image classifier to determine the level of difficulty associated with geolocating an image. This study designs, implements, and evaluates an automated image classifier that detects those images that are “easy” to geolocate. This study answers the following research question:

**Can a classifier automatically predict the difficulty of geolocating an image?**

### III. DATASET, METHODOLOGY AND EXPERIMENT

In this section, we first give an overview of the dataset and the methods employed in the experiment. Following that, we provide a detailed explanation of the experimental design.

To answer the research question, we devised two major tasks: First, images taken from a reference dataset are shown to a group of domain experts that label each with a grade indicating the level of geolocalization. The second task involves the training a deep learning model using the data annotated in the previous task.

| Records      | Catania floods |      | EU floods |     | Haiti earthquake |      | Croatia earthquake |      | All datasets |      |
|--------------|----------------|------|-----------|-----|------------------|------|--------------------|------|--------------|------|
|              | N              | %    | N         | %   | N                | %    | N                  | %    | N            | %    |
| Geolocatable | 113            | 100  | 271       | 100 | 1,462            | 100  | 627                | 100  | 2,473        | 100  |
| Geolocated   | 70             | 61.9 | 214       | 79  | 961              | 65.7 | 549                | 87.6 | 1,794        | 72.5 |
| Timeout      | 43             | 38.1 | 57        | 21  | 501              | 34.3 | 78                 | 12.4 | 679          | 27.5 |

TABLE I  
SUMMARY OF THE IMAGE GEOLOCATING RESULTS CONSIDERING JUST THE TOTAL NUMBER OF GEO-LOCATABLE RECORDS

#### A. Dataset

For our reference dataset, we utilized images that were geolocated by GIS (Geographic Information System) experts. These images were gathered as part of EMSV070 - Reference Datasets, which focused on creating the reference dataset for SM (Social Media) Image Geolocation in DRM (Disaster Risk Management), as requested under the Copernicus Emergency Management Services framework. The dataset includes geolocated images and text gathered during four different disasters: (1) Croatia Earthquake, 2020 (30.7%); (2) EU Floods, 2021 (3.9%); (3) Central European floods, 2021 (11.9%); and (4) Haiti Earthquake, 2010 (53.5%) (see Table I).

Preferably, the damage in the image (flooded point, damaged building, etc.) was selected for geolocation. In total, we had 5,430 records of images. From 5,430 records, there were 693 repeated and 4,737 analyzed. The details of 4,737 records were 2,264 non-geolocatable and 2,473 geolocatable. Out of 2,473 geolocatable records, we opted for 1,794 records, while 679 records were excluded because the images are inaccessible.

The experts used Google Street View, Mapillary, Google Earth Pro, and Google Maps for localizing the places, recognizing the visible elements in the images, and/or for the subsequent image geolocation. Additionally, the already recognized images during the execution of the task were also considered for the geolocation of new images.

Consequently, this study uses images that experts have geolocated at least at the city level, and in 90% of cases, with a precision ranging from 10 to 100 meters.

Experts in the domain of Disaster Response, specifically the European network VOST Portugal (VOSTPT)<sup>3</sup> classified the images of the reference dataset, each based on the difficulty of geolocating them. Four different levels were utilized: (1) Easy, (2) Medium, (3) Difficult, or (4) Impossible to geolocate (see Figure 1). Each image was labeled at least by five different raters.

The crowdsourced analysis of the images was carried out through the Pybossa framework [10]. More precisely, this task was executed by using an instance of Pybossa hosted at the Citizen Science Center Zurich<sup>4</sup>, a joint initiative created by the University of Zurich and ETH Zurich. The name of the application is Citizen Science Project Builder (CSPB). In this task, the crowd determined the level of difficulty in

<sup>3</sup><https://vosteuropa.org/>

<sup>4</sup><https://citizenscience.ch/en/>



Fig. 1. Examples of geolocation difficulty levels for images

geolocating each image by assigning four labels, i.e., easy, medium, difficult, and impossible.

Regarding the dataset used for the second task, i.e., training our deep learning model, we adopted Majority Voting and Dawid-Skene [11] consensus mechanisms to aggregate the annotations performed by the VOST community. We elaborate both mechanisms in Section IV.

#### B. Training a CNN to predict the level of difficulty in geolocating a given photo

To predict the level of difficulty in geolocating a photo, this paper proposes the use of a Convolutional Neural Network (CNN). The training the CNN is performed through the following steps:

Initially, we imported images from the annotated training dataset. We loaded images and distributed them among the classes “Easy” and “Difficult”. Then, we used all the labels existing in the annotations, but after some preliminary analysis, we considered reducing the complexity of the classification. Hence, the dataset used to train the model included only the images whose consensus labels were Easy, Difficult, or Impossible, making the classification a binary task classifying between Easy (Easy) and Difficult (Difficult or Impossible). The rationale behind this choice was to focus on accurately identifying True Positive values (the Easy class).

We then deduplicated the data using hashing algorithms<sup>5</sup> that are particularly good at finding exact duplicates, as well as CNNs, which are also good at finding quasi-duplicates. Subsequently, we employed standard data analysis techniques to examine the distribution of images.

We divided the data set into training, validation, and test sub-datasets with ratios of 70% (448 images), 20% (150 images), and 10% (61 images), respectively. During data preparation phase, we performed data augmentation and pre-processing. The data augmentation enhanced the diversity of our training set by incorporating random yet realistic, image rotation and image flipping. This step basically multiplies the quantity of images used in our training. The number of initial epochs was set to 25 (and also 25 for the fine-tuning)

<sup>5</sup><https://pypi.org/project/imagededup/>

In the experiment, we chose the EfficientNet-B6 due to its good balance between computational demand and model performance [12]. Additionally, it aligns with the image size commonly found on social media, at a resolution 528x528 pixels. Recognizing that training a deep learning model from scratch on a small dataset might lead to suboptimal results, we decided to employ transfer learning. This process entailed taking a pre-existing model that was trained on a large dataset and adapting it to our task. This approach not only reduces the amount of required data but also shortens the training time while enhancing the model’s performance when dealing with smaller datasets.

To adjust the pre-trained model for our particular task, we conducted fine-tuning on the last 66 layers of the EfficientNet-B6 in a two-step process : first freezing all layers and training only the top layers, and then unfreezing 66 layers (out of 666) and fitting the model using a smaller learning rate. A relatively large learning rate (1e-2) was used for the first step. Fine-tuning encompasses adapting the weights of a pre-trained model to new data. This approach enables us to make use of both the general feature extraction learned during pre-training and the specific patterns found in our disaster imagery.

To conclude, the experiment aimed to leverage the power of EfficientNet-B6 and the robustness of Noisy Neighbour pre-training by fine-tuning the model to our specific classification. Figure 2 shows a comparison of accuracy while testing different weights.

The base classifiers were trained with the objective of decreasing the cross-entropy loss. Cross-entropy measures how well a classifier approximates the probabilities of its predictions.

## IV. EXPERIMENTAL RESULTS

This section explains the experimental results from the different tasks presented in Section III. The goal of this section is to answer the research question proposed in Section II.

As the pilot study, we initially worked with 1,794 photos annotated by the VOSTPT community using the CSPB tool<sup>6</sup>, with the five annotator for each image. As the total, about 8,970 classifications were performed by the VOSTPT community. Each image was classified by the level of geolocation difficulty, i.e. Easy, Medium, Difficult, and Impossible. Using the annotation dataset that we had, we assessed the inter-rater agreement. Fleiss’ Kappa<sup>7</sup> is a common statistical measure to evaluate the reliability of agreement among raters when it comes to categorical ratings. We calculated the Fleiss’ Kappa for the VOSTPT dataset using the Python library Crowdanalysis [13]. The value was 0.288 which indicates a Fair agreement on the kappa scale among the VOSTPT community for this dataset. The low value indicates that determining the level of difficulty in assigning a geolocation to a social media image is a challenging task and the annotation result might be noisy.

To create the dataset for the next step, which involves training the CNN, we had to reach a consensus among all five

<sup>6</sup><https://lab.citizenscience.ch/en/tools/projectbuilder>

<sup>7</sup>[https://en.wikipedia.org/wiki/Fleiss\\_kappa](https://en.wikipedia.org/wiki/Fleiss_kappa)



Fig. 2. ImageNet (left) and Noisy student (right) weight performance comparison on model training

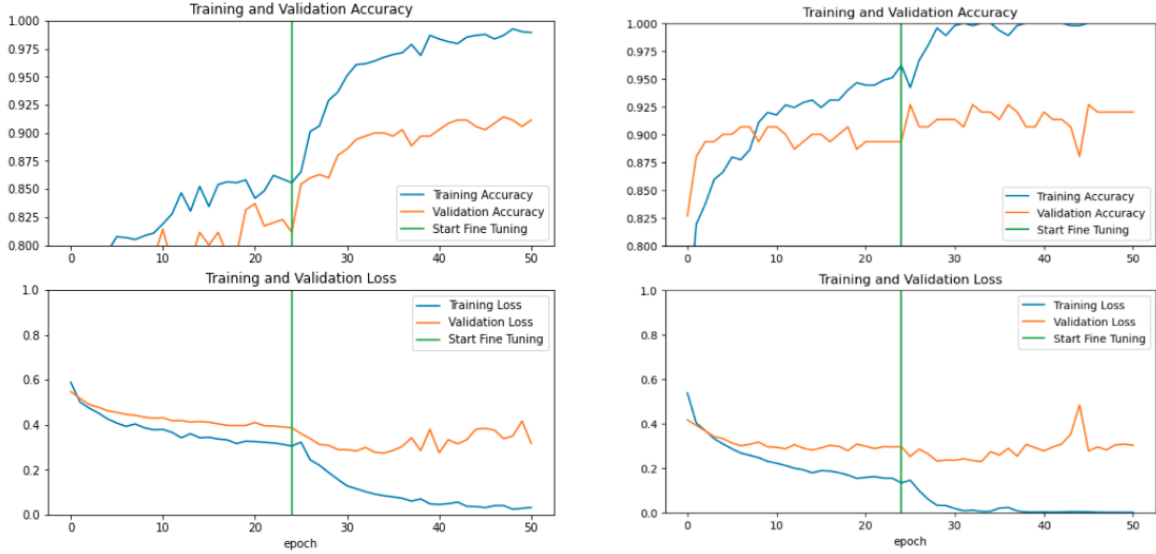


Fig. 3. Comparison of model training performances between Majority Voting (left) and Dawid-Skene (right) consensus training datasets

annotators for each image. The Crowdfunder library, besides the standard Majority Voting, provides advanced probabilistic methods of consensus [14] such as the seminal Dawid-Skene model (DS). The DS method enables modelling individual annotator behaviour, thus with enough data, it yields more reliable consensus results which can prove crucial in disaster management scenarios, as mentioned earlier.

Table II shows the percentage of warnings for each consensus result. The warning was a condition where there was only a  $< 0.1$  difference between the probabilities of the top and second-best estimated consensus classes for the photo. The values in bold indicate the lower values between models trained with DS and MV consensus datasets.

We ran two experiments with two data sets originating from

|           | Easy         | Medium       | Difficult   | Impossible  | Not Relevant |
|-----------|--------------|--------------|-------------|-------------|--------------|
| <b>DS</b> | <b>4.4%</b>  | <b>5.2%</b>  | <b>9.7%</b> | <b>0.7%</b> | 0.43%        |
| <b>MV</b> | <b>57.2%</b> | <b>58.8%</b> | <b>10%</b>  | <b>2.7%</b> | <b>0%</b>    |

TABLE II  
PERCENTAGE OF CONSENSUS RESULT WARNINGS. DS REPRESENTS DAWID-SKENE, WHILE MV STANDS FOR MAJORITY VOTING

Majority Voting (MV) and Dawid-Skene (DS) consensus algorithms and compared the performances of the two models. The model trained with the DS dataset presented better results than the one with Majority Voting, as we expected since DS weighs the raters' answers based on their error rates computed by the same algorithm. That is, the



Fig. 4. Examples of correctly classified images by the model. The text in the upper of each image corresponds to image geolocation difficulty level

annotations of the raters who perform better (i.e., make fewer errors) have more influence on the consensus calculation. Thus, we expect to have a more reliable consensus with DS. Figure 3 demonstrates the comparison of the training accuracy with MV and DS datasets and their evolution after being fine-tuned.

This initial result indicates that classifiers have the ability to estimate the challenge of geolocating a social media image. This step is crucial because it helps to sort out images that are impossible to geolocate and it allow us to improve the ratio of images geolocated per person right after a disaster happens when the time is crucial. Through this approach, easy images would be scheduled with higher priority than those difficult to geolocate, saving invaluable time for disaster management. We assess the validity of our hypothesis in the following section. Figure 4 shows examples of classification correctly inferred by the model.

Table III demonstrates the confusion matrix result of evaluating the model trained with the DS dataset. The model presents an accuracy of 0.87, a precision of 0.88, and a recall of 0.88. Assessing the difficulty of geolocating an image remains a challenging task even for the experts, as indicated by the Kappa value mentioned earlier. However, the evaluation of the proposed model indicates that it is possible to automatically predict the difficulty of geolocating an image. This result answers our research question.

|           | Predicted |      |
|-----------|-----------|------|
| Actual    | Difficult | Easy |
| Difficult | 29        | 4    |
| Easy      | 4         | 24   |

TABLE III

CONFUSION MATRIX FOR THE CNN MODEL IN BINARY CLASSIFICATION

## V. CONCLUSIONS AND FUTURE WORK

Finding the location of social media images related to a disaster is a key process to make the data actionable, i.e., data that can be used to enable better-informed and timely decisions. The task of geolocating social media images is commonly carried out by humanitarian communities, including Stand By Task Force, GISCorp, or VOSTPT. Crowdsourcing the location of social media images is commonly recognized as a challenging task. This difficulty often delays the assessment of the impact of natural or man-made disasters right after they occur. Our findings demonstrated that it is indeed possible to automatically predict the difficulty of geolocating an image. The suggested method could enable prioritizing images that

are “easy” to geolocate by the crowd and thus could lead to a quicker and more effective analysis of social media data.

As future work, the proposed approach will be put to the test in a real-time crisis response to demonstrate the potential of using social media data to assess the impact of a natural or man-made disaster within 24 hours.

## ACKNOWLEDGMENT

This work has received funding from the Indonesia Endowment Fund for Education (LPDP) and the European Union’s Horizon 2020 research and innovation programme K-HealthInAIR project with ID 101057693.

## REFERENCES

- [1] B. Huang and K. M. Carley, “A large-scale empirical study of geotagging behavior on twitter,” in *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ser. ASONAM ’19. New York, NY, USA: Association for Computing Machinery, 2020, p. 365–373.
- [2] F. Murgese, G. Costa-Alcaina, M. O. Mülâyim, J. Cerquides, and J. L. Fernandez-Marquez, “Automatic outdoor image geolocation with focal modulation networks,” in *Artificial Intelligence Research and Development: Proceedings of the 24th International Conference of the Catalan Association for Artificial Intelligence*. IOS Press, 2022.
- [3] G. Scalia, C. Francalanci, and B. Pernici, “CIME: Context-aware geolocation of emergency-related posts,” *Geoinformatica*, vol. 26, no. 1, pp. 125–157, 2022.
- [4] P. Rufolo, D. Muraro, and V. Lorini, *Social media image analysis in the immediate aftermath of the 2020 Beirut blast*, 2021.
- [5] V. Lorini, E. Panizio, and C. Castillo, “Smdrm: A platform to analyze social media for disaster risk management in near real time,” in *In Workshop Proceedings of the 16th International AAAI Conference on Web and Social Media*, 2022.
- [6] C. Bono, B. Pernici, J. L. Fernandez-Marquez, A. R. Shankar, M. O. Mülâyim, and E. Nemni, “TriggerCit: Early Flood Alerting using Twitter and Geolocation—a comparison with alternative sources,” in *Proc. ISCRAM 2022, Tarbes, France*, May 2022.
- [7] V. Lorini, P. Rufolo, and C. Castillo, “Venice was flooding ... one tweet at a time,” *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. CSCW2, nov 2022. [Online]. Available: <https://doi.org/10.1145/3555107>
- [8] H. B. Firmansyah, J. L. Fernandez-Marquez, J. Cerquides, V. Lorini, C. A. Bono, and B. Pernici, “Enhancing disaster response with automated text information extraction from social media images,” in *The 9th IEEE International Conference on Big Data Computing Service and Machine Learning Applications*, 2023.
- [9] A. Ravi Shankar, J. L. Fernandez-Marquez, B. Pernici, G. Scalia, M. R. Mondardini, and G. Serugendo, “Crowd4EMS: A crowdsourcing platform for gathering and geolocating social media content in disaster response,” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-3/W8, p. 331–340, 2019.
- [10] D. L. González, alejandrodob, therealmarv, M. Keegan, A. Mendes, R. Pollock, N. Babu, F. Fiordalisi, N. A. Oliveira, K. Andersson, and et al., “Scifabric/pybossa: v3.1.3,” Jun 2020.
- [11] A. P. Dawid and A. M. Skene, “Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm,” *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 20–28, 1979, publisher: [Wiley, Royal Statistical Society]. [Online]. Available: <https://www.jstor.org/stable/2346806>
- [12] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” 2020.
- [13] J. Cerquides and M. O. Mülâyim, “crowdanalysis: A software library to help analyze crowdsourcing results,” <https://doi.org/10.5281/zenodo.5898579>, Jan. 2022.
- [14] J. Cerquides, M. O. Mülâyim, J. Hernández-González, A. Ravi Shankar, and J. L. Fernandez-Marquez, “A conceptual probabilistic framework for annotation aggregation of citizen science data,” *Mathematics*, vol. 9, no. 8, 2021. [Online]. Available: <https://www.mdpi.com/2227-7390/9/8/875>