Evaluating the impact of floods on gender equality from social media evidence

Olimpia Rivera, Juan Felipe Calderon
Politecnico di Milano

Paul Planchon
Politecnico di Milano & CyTech Cergy

Barbara Pernici
Politecnico di Milano
barbara.pernici@polimi.it

ABSTRACT

Climate change is one of the major challenges of our times. As a matter of fact, UN Sustainable Development Goal 13 (SDG 13) is about climate action and is one of the 17 Sustainable Development Goals established by the United Nations in 2015. It aims to "take urgent action to combat climate change and its impact". It is widely recognized that climate change does not affect people equally and often, women are more vulnerable to climate change than men. This paper aims at investigating whether relevant insights related to the relationship between climate change (SDG 13) and gender equality (SDG 5) could be extracted from social media, in particular from Twitter, by exploiting image recognition technologies and crowdsourcing techniques.

CCS CONCEPTS

• **Information systems** → **Social tagging systems**; **Data cleaning**; **Spatial-temporal systems**.

KEYWORDS

citizen science, SDG indicators, social media image analysis, floods, gender equality

ACM Reference Format:
Olimpia Rivera, Juan Felipe Calderon, Paul Planchon, and Barbara Pernici. xxx. Evaluating the impact of floods on gender equality from social media evidence. In *xxx*. ACM, New York, NY, USA, 5 pages. https://doi.org/xxx

1 INTRODUCTION

It is widely recognized that climate change does not affect people equally and often, women are more vulnerable to climate change than men. The related disasters and impacts often intensify existing inequalities, vulnerabilities, economic poverty and unequal power relations. For these reasons, the measurement of progress on SDG 13 needs also to assess how climate change affects women and whether climate change responses are properly taking into account the existing inequalities. A synthetic representation of the evaluation of the progress of SDGs and gender equality is shown in [4]. Social media represent nowadays a powerful tool to extract information that, if properly aggregated and analyzed can provide relevant insights and indicators. The urgent need to assess whether and how climate change affect women, requires a shift from readily available data to more 'difficult-to-measure' indicators. The goal of this paper is to present a methodological approach in which both AI and crowdsourcing have been used to process images from Twitter posts with the goal of getting insights from visual evidence on the factors related to the gender differentiated impacts of these type of climate-related disasters. The paper provides the first results derived from the ongoing analyses.

The paper has the following structure. In Section 2 we discuss the state of the art. In Section 3 we describe the approach and methodology followed to extract flood images from Twitter. Finally in Section 4 we illustrate and evaluation the resulting dataset and the related conclusions that can be drawn.

xxx, xxx, xxx

ACM ISBN xxx…$xxx
https://doi.org/xxx

2 RELATED WORK

Several issues have been identified as keys in understanding why women prove to be more vulnerable to climate-related disasters (floods in particular) than men [6, 10, 12]:

Poverty: women face higher risk in situations brought by climate changes because they make up the majority of the world's economically poor. 70% of the world's poor are indeed women.

Inequities in labour division: Women play a critical role in agricultural and pastoral livelihoods, often bearing significant responsibility for managing critical productive resources such as land, water, livestock, biodiversity, fodder, fuel, and food. They are often responsible for gathering and producing food, collecting water and sourcing fuel for heating and cooking. With climate change, these tasks are becoming more difficult as distances travelled by women to access natural resources (such as water, fuel wood, fodder, food, pastures, medicinal plants, fuel, and crops) increase. Climate change-induced flooding is over time likely to increase women's workloads in domestic fuel and water collection in some regions. This will therefore, reduce their time available for childcare, education and participation in public life. They also carry out a disproportional amount of daily labour compared to men in household and community spheres, such as cooking, cleaning, child care, care of older or sick family members. The unequal gender division of labour is further skewed by climate change, as often, when livelihoods are destroyed and productive assets are eroded, men tend to migrate out in search of income generating opportunities. This in-turn intensifies women's workloads as they struggle to add the work that men used to manage to their own daily workloads.

Violence: Not only do women constitute the majority of victims of floods and experience the greatest difficulty in recovering from a disaster, but they are also more likely to be subjected to sexual violence in the aftermath of disasters. Numerous reports have documented an increase in such violence against women following environmental disasters. In addiction, inadequate temporary accommodation in post-disaster situations renders women vulnerable to sexual and gender-based violence.

Other critical gender inequalities: As previously stated, women are disproportionately vulnerable to climate change, but this is not because there is something inherently vulnerable about women, but because of socio-cultural structures that deprive women of access to resources, decision-making, information, education, etc. Just to mention few examples: in countries where gender inequality is more severe, death rates for women in climate-related disasters are significantly higher. The reason for this could simply be that women are not taught to swim. In cultures that restrict women from leaving their houses unaccompanied or from learning to swim or to climb trees, women may suffer greater injury and fatality in some kinds of climate change-induced natural disasters. There is also the fact that women in more unequal societies do not tend to move about in public spaces, which means they will not hear warnings, and are unable to get out fast enough. Lastly, women are more likely to have dependents, such as children and elderly or sick relatives, which affects their ability to leave the damaged area.

The use of Citizen Science for collecting information relevant for SDGs is discussed in [5]. Among the possible sources of information in this context social media and crowdsourcing are being considered. One of the challenges of Citizen Science is the quality of extracted information, in particular due to the variety of the background of the contributors and possibly open and anonymous contributions. Strategies for improving the information quality, such as selecting contributors with a given background, or training the contributors, are illustrated, and also human factors, such as bias, are discussed in [8].

Social media analysis in emergencies, and in particular in floods and earthquakes, has been discussed in [7]. The paper describes the challenge of selecting the relevant posts and defines several technologies, such as crowdsourcing and automatic annotations of posts with ML trained models, for improving the quality of the results.

The difficulty of retrieving relevant posts is discussed in [1], in which an approach to automatically improving the search keywords during an emergency event is proposed, and in [2] to provide a selection of good quality posts and images for crowdsourcing, filtering the posts on the basis of the characteristics of the images.

In [11] the authors discuss social media analysis based on Twitter towards understanding the societal impact of hydrometeorological events. The analysis focuses on analyzing the text of the tweets, for understanding their intention and the source of the post. In particular, most of the tweets are descriptive or informative. The study also puts forward the difficulty in the analysis of datasets about such events, with the need of a manual annotation of all the tweets to identify the relevant ones.

In the VisualCit approach [9] developed in the Crowd4SDG[1] project the focus is on analyzing images from tweets automatically selecting the relevant tweets with ML classifiers, filtering out images which are not photos, selecting images in public places, and automatically geolocating them. Crowdsourcing to gather additional information for tweets annotation is performed only on the selected tweets, significantly reducing the number of tweets to be examined. The potential of deriving statistical indicators from the collect knowledge is also discussed based on a case study.

In the present paper we show how to develop VisualCit further to collect posts about floods and identify gender-related problems.

3 BUILDING A DATASET ON FLOODS AND GENDER EQUALITY

In this paper, we discuss the construction of a dataset for analyzing the impact of floods on gender equality. The goal is to be able to gather visual evidence on critical issues from images in tweets.

3.1 Methodology

The following requirements and constraints have been defined: i) the dataset should be based on images extracted from social media (in particular Twitter is considered in the study); ii) crowdsoucing by citizen scientists should be used for evaluating both the context of the post and the relevance to the research question. The size and the composition of the available crowd has proven one of the challenges in previous social media analysis with crowdsourcing projects. Therefore a requirement is iii) to collect an initial dataset to be analyzed by the crowd mainly composed of relevant posts, limiting the number of posts sent to crowdsourcing.

As discussed in Section 2, one of the challenges is to select the search keywords for initial selection of posts and to define the questions to be asked to the crowd, in order to gather evidence from the collected images.

3.2 Tool environment

The developed tool environment is based on VisualCit [9], which has the goal of improving the quality of the initial dataset. In this work, we used the VisualCit components that allow filtering images to retain only non-duplicate images which are photos in public places, and safe for work during the crowdsourcing phase.

In addition, for the purpose of this project, two components were specifically designed for collecting data about gender equality during floods. First, as we are focusing on getting evidence from the field during a flood, we developed a flood classifier to retain only images related to flooded areas (see Section 3.2.1, then we designed a crowdsourcing project to investigate about inequalities in this context, as described in Section 3.2.2.

3.2.1 Flood classifier. As previously stated, for the purpose of this work, it has been decided to focus on flood and therefore a flood classifier module has been developed and added to the pipeline described above. To perform image classification we used Deep Learning techniques and in particular Convolutional Neural Networks that are the de facto standard techniques for solving image classification tasks. We tried different models such as ResNet50, MobileNet, and Xception all having similar performances. Xception [3] was the model selected for the classifier. We started from the pre-trained network on ImageNet and we applied fine tuning using the dataset we created. In particular we started from public datasets such as the eu-flood[2] and other sources[3]. For the non flood images

[1] https://crowd4sdg.eu/

[2] https://github.com/cvjena/eu-flood-dataset

[3] https://www.dropbox.com/sh/grxeep1k9a0yziq/AAByrZYB-jGQoTvb0Yp22fJFa

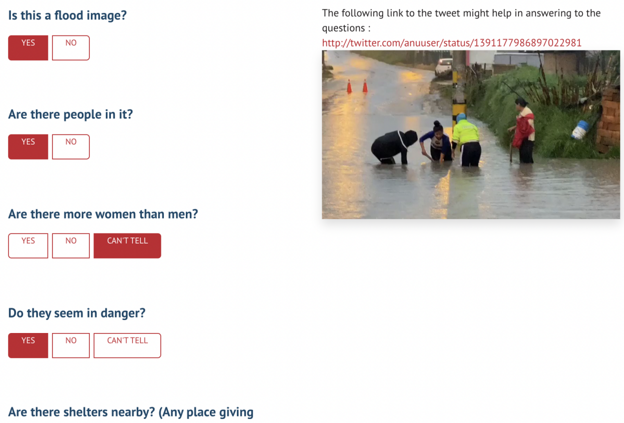

Figure 1: Crowdsourcing interface

we extracted some images from the COCO dataset and we crawled some random images using the VisualCit pipeline. The validation accuracy was around 0.96% but after testing it with external source images we found out that it was not working as expected. Specifically, we noticed that the model did not recognize images from rural areas (that are the most common in Africa and in underdeveloped countries). The reason lies in the fact that the dataset was composed mostly of European flood images. Therefore, we added African floods images extracted from Google. To further improve the classifier we decided to use a cyclic approach:

(1) Use the VisualCit pipeline and the flood filter to gather new flood images from a new crawling.
(2) Manually label the predicted flood images retrieved from the previous step.
(3) Add the manually labeled images to the dataset.
(4) Train flood filter with the larger dataset.
(5) Go back to step 1

We completed the training cycle when precision of the classifier reached 0.95.

3.2.2 Crowdsourcing. The crowdsourcing application leverages the crowd in order to extract relevant information from the resulting pictures. We used the PyBossa-based Project Builder tool from Citizen Science Center Zurich[4], in which the selected images are shown to the crowd workers, together with questions regarding the depicted people, the activities they are doing and the environment they are in. Unique tasks are created for each tweet and tasks are considered completed once three different crowd workers analyze the tweet.Figure 1 shows the provided interface and som e of the questions. The project can be accessed at https://lab.citizenscience.ch/en/project/329. The results of all the questions will be discussed in Section 4.

4 ANALYSIS OF THE RESULTS

Once an overall progress of 70% was reached (317 completed tasks over 452), the data were analyzed using matplotlib.pyplot. For the completed tasks (analyzed by at least three different crowd workers), the answers selected by the majority of the workers were considered for the analysis.

Some interesting results were obtained by analyzing pictures in which more women are present.

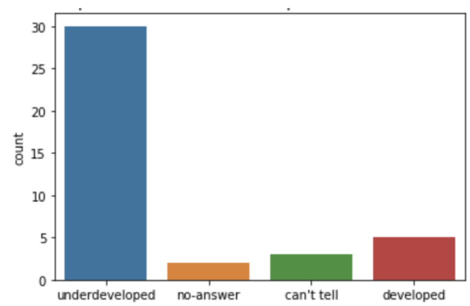

Figure 2: "Do you think the photo was taken in a developed or in an underdeveloped country?"

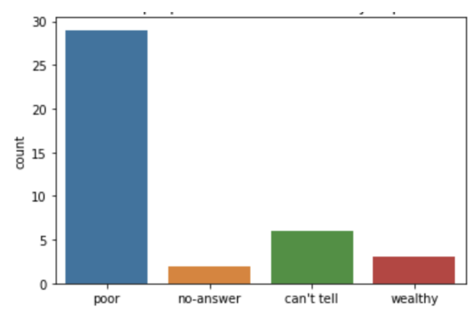

Figure 3: "Do the people involved seem wealthy or poor?"

As shown in the histograms in Figures 2 and 3, those pictures are depicting mainly poor people and in the majority were taken in underdeveloped countries, according to almost all of the crowd workers' answers. Those results have a correlation with what has been stated above regarding poverty. Poverty and vulnerability are not uniformly correlated, but economically poor people and socially excluded groups tend to suffer disproportionately from vulnerability and the majority of world's poor are, indeed, women.

Furthermore, the histograms in Figure show a comparison between the answers to the question "Do they seem in danger?" associated to pictures in which more men are depicted and the answers to the same question associated to pictures with more women. Interestingly enough, men are apparently exposed to less danger in floods situations with respect to women.

Similar results were obtained by analyzing the data related to shelters: as shown in Figure 5, the percentage of pictures showing shelters (any place giving temporary protection from bad weather) and in which mostly women are present is lower with respect to the percentage of pictures showing shelters and in which mostly men are present.

It is important to remark, however, that these considerations cannot be considered particularly statistically meaningful because

[4]https://citizenscience.ch/en/

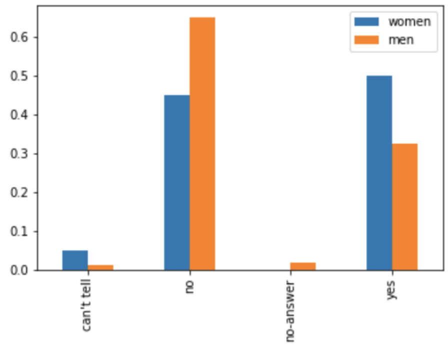

Figure 4: "Do they seem in danger?"

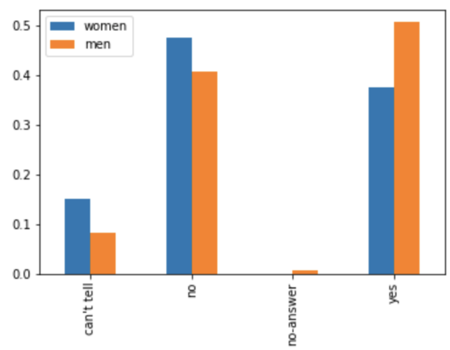

Figure 5: "Are there shelters nearby?"

of the following reasons: limited number of available data, relatively few images depicting more women than men with respect to the ones depicting more men than women and, not exact knowledge of actual number of women and men in the images.

Finally, we analyzed the data related to the images' locations and combined it with the OECD (Organization for Economic Co-operation and Development) "Gender, Institutions and Development" Database. As shown in Figure 6, Latin American and African countries seem to be the ones where the majority of the floods shown in the images hit. This type of result is strictly correlated to the outcome obtained analyzing the OECD Database[5], which shows that legal frameworks do not protect women from violence in many of the most frequently hit countries.

Lastly, we analyzed the labels collected through crowdsourcing to characterize the scene of the photo, we classified them manually by topic to visualize and compare the most recurrent topics present in the answers to the question regarding the activities that people in the images are doing. It is interesting to note that "escaping" and another topic associated to movement ("walking in water") are the most frequently associated to pictures in which more women are present. Topics such as "rescuing" and "cleaning" instead, which might come within the scope of "repairing the post-disaster situation", are only present in pictures in which mostly men are present. (Figure 7).

[5] https://stats.oecd.org/viewhtml.aspx?datasetcode=GIDDB2019

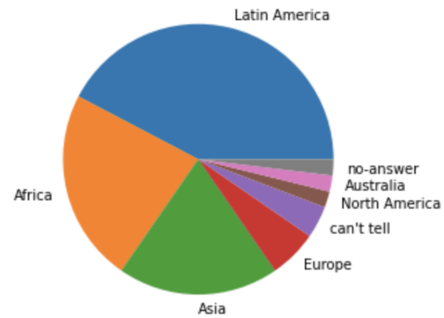

Figure 6: "In which of the following regions do you think the photo was taken?"

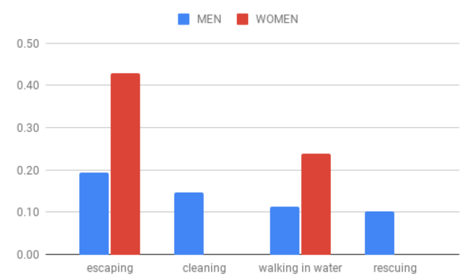

Figure 7: "Most popular topics by gender"

5 CONCLUDING REMARKS

The paper presents a case study in which evidence of gender inequality is obtained analyzing posts from Twitter and using Machine Learning to reduced the number of irrelevant posts for the research, followed by crowdsourcing. The initial results show a clear evidence that the number of women in danger is higher than that of men. The case study will be continued to increase the number of posts for increasing the gathered evidence and to analyze the differences emerging in areas of the world.

ACKNOWLEDGMENTS

This work was partially funded by the European Commission H2020 project Crowd4SDG "Citizen Science for Monitoring Climate Impacts and Achieving Climate Resilience", project no. 872944. The authors thank the Citizen Science Center (https://citizenscience.ch/en/) for providing the Project Builder PyBossa software platform.

REFERENCES

[1] Andrea Autelitano, Barbara Pernici, and Gabriele Scalia. 2019. Spatio-temporal mining of keywords for social media cross-social crawling of emergency events. *GeoInformatica* 23, 3 (2019), 425–447. https://doi.org/10.1007/s10707-019-00354-1

[2] Sara Barozzi, Jose Luis Fernandez-Marquez, Amudha Ravi Shankar, and Barbara Pernici. 2019. Filtering images extracted from social media in the response phase of emergency events. In *Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management, València, Spain, May 19-22, 2019*, Zeno Franco, José J. González, and José H. Canós (Eds.). ISCRAM Association. http://idl.iscram.org/files/sarabarozzi/2019/1881_SaraBarozzi_etal2019.pdf

[3] François Chollet. 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1251–1258.

[4] Equal measures. link checked 21/5/2021. Equal Measures 2030. https://data.em2030.org/goals/sdg13/

[5] Steffen Fritz, Linda See, Tyler Carlson, Mordechai Muki Haklay, Jessie L Oliver, Dilek Fraisl, Rosy Mondardini, Martin Brocklehurst, Lea A Shanley, Sven Schade, et al. 2019. Citizen science and the United Nations sustainable development goals. *Nature Sustainability* 2, 10 (2019), 922–930.

[6] Kathleen Grantham. 2020. Mapping gender data gaps: an SDG era update.

[7] Clemens Havas, Bernd Resch, Chiara Francalanci, Barbara Pernici, Gabriele Scalia, Jose Luis Fernandez-Marquez, Tim Van Achte, Gunter Zeug, Maria Rosa (Rosy) Mondardini, Domenico Grandoni, Birgit Kirsch, Milan Kalas, Valerio Lorini, and Stefan Rüping. 2017. E2mC: Improving Emergency Management Service Practice through Social Media and Crowdsourcing Analysis in Near Real Time. *Sensors* 17, 12 (2017), 2766. https://doi.org/10.3390/s17122766

[8] Roman Lukyanenko, Andrea Wiggins, and Holly K. Rosser. 2020. Citizen Science: An Information Quality Research Frontier. *Inf. Syst. Frontiers* 22, 4 (2020), 961–983. https://doi.org/10.1007/s10796-019-09915-z

[9] V. Negri, D. Scuratti, S. Agresti, D. Rooein, G. Scalia, J.L. Fernandez-Marquez, A. Ravi Shankar, M. Carman, and B. Pernici. 2021. Image-based Social Sensing: Combining AI and the Crowd to Mine Policy-Adherence Indicators from Twitter. In *ICSE, Track Software Engineering in Society*.

[10] Christian Nellemann, Ritu Verma, and Lawrence Hislop. 2011. Women at the frontline of climate change: Gender risks and hopes. *A Rapid Response Assessment. United Nations Environment Programme, GRID-Arendal* (2011).

[11] Esther Oliver, María Carmen Llasat, Montserrat Llasat-Botija, and Javier Díez-Palomar. 2021. Twitter's Messages about Hydrometeorological Events. A Study on the Social Impact of Climate Change. *Sustainability* 13, 6 (2021), 3579.

[12] Cecilia Sorensen, Virginia Murray, Jay Lemery, and John Balbus. 2018. Climate change and women's health: Impacts and policy directions. *PLoS Medicine* 15, 7 (2018), e1002603.