Emerging Issues Report

K4D | Knowledge, evidence and learning for development

Predictive analytics in humanitarian action: a preliminary mapping and analysis

Kevin Hernandez
Institute of Development Studies

Tony Roberts
Institute of Development Studies

June 2020

About this report

The K4D Emerging Issues report series highlights research and emerging evidence to policy-makers to help inform policies that are more resilient to the future. K4D staff researchers work with thematic experts and the UK Government's Department for International Development (DFID) to identify where new or emerging research can inform and influence policy.

This report is based on 14 days of desk-based research in May 2020.

K4D services are provided by a consortium of leading organisations working in international development, led by the Institute of Development Studies (IDS), with the Education Development Trust, Itad, University of Leeds Nuffield Centre for International Health and Development, Liverpool School of Tropical Medicine (LSTM), University of Birmingham International Development Department (IDD) and the University of Manchester Humanitarian and Conflict Response Institute (HCRI).

For any enquiries, please contact helpdesk@k4d.info.

Acknowledgements

We would like to thank Nicholas Leader, who provided advice and guidance on the development of this report and David Fallows, who served as an external expert reviewer.

We would also like to thank Alice Shaw and Lewis Small, who copyedited this report.

Suggested citation

Hernandez, K. and Roberts, T. (2020). *Predictive Analytics in Humanitarian Action: a preliminary mapping and analysis.* K4D Emerging Issues Report 33. Brighton, UK: Institute of Development Studies.

UKaid
from the British people

Contents

1. Executive summary

Humanitarian predictive analytics is the use of big data to feed machine learning and statistical models to calculate the probable characteristics of humanitarian emergencies. The technology is being used to forecast the likely trajectory and features of humanitarian emergencies including pandemics, famines, natural disasters and refugee movements. This form of artificial intelligence is used to predict where and when disasters will unfold, what the defining characteristics of the situation will be and who will be the most affected populations. Accurate advance prediction enables the pre-positioning of emergency relief finance, supplies and personnel.

Forecasting and early warning systems have always been a component of humanitarian action. However, the rapid expansion of computing power and big data has dramatically increased the potential for predictive analytics in evermore areas of humanitarian action. In the last few years, the term predictive analytics has come to refer primarily to a digital process, drawing on multiple sources of electronic data feeding machine learning algorithms to inform statistical models that compute the probability of different humanitarian outcomes. Historic data of previous humanitarian events plus mobile phone records and social media posts can provide the high volumes of data needed to analyse food security, predict malnutrition and inform aid deployment. Satellite images, meteorological data and financial transactions can be used to track and predict the escalation and trajectory of refugee movements.

This rapid review research provides the most comprehensive mapping and analysis of predictive analytic initiatives in humanitarian aid to date. It documents 49 projects including a variety of novel applications (see Appendix for details). It provides a typology of predictive analytics in digital humanitarianism and answers a series of key questions about patterns of current use, ethical risks and future directions in the application of predictive analytics by humanitarian actors.

The study took 14 days in May 2020. Forty-nine predictive analytics projects were mapped and analysed according to the main phases of the humanitarian cycle, type of predictions made, sector of application, geography of application, and technical approach used. Despite the limitations of rapid response research, some preliminary recommendations are made on the basis of the findings listed below.

Main findings:

- Our research shows that predictive analytics is being used in the mitigation, preparedness and response phases of the humanitarian lifecycle, but not the recovery phase.
- Predictive analytics is also used by humanitarian agencies for functions such as human resource management, fundraising and logistics.
- Predictive analytics is most often used by projects covered in this review to predict **where** humanitarian crises will occur (71% of initiatives) and **who** will be affected (40%). Less often it is used to predict **what** the affected situation will look like (26%) and **when** events will occur (18%).
- *By sector,* predictive analytics is being applied in a wide variety of humanitarian applications. The most common are prediction of **disease outbreak** (9 initiatives), **migration** (9), **conflict** (7), **disaster risk reduction** (6), and **food security** (4).

- *Geographically*, the initiatives covered in the review were primarily in Africa and the Arab world, with fewer applications in Asia and Latin America.
- *Technically,* most initiatives use historical humanitarian data combined with machine learning and statistical modelling to produce predictions.
- The study details examples of a wide range of data sources, data collection techniques, machine learning and analytical models.
- Predictive analytics is currently being used to complement rather than to replace traditional humanitarian analysis and forecasting.
- Humanitarian predictive analytics is being used most often by large international agencies and small start-up companies.
- We found little evidence of affected populations playing a significant role in the design or management of predictive analytics in humanitarian work.
- Almost half of the initiatives (23) claimed that predictive analytics would improve efficiency by saving time or money although we were unable to validate these claims.
- The ecosystem of humanitarian predictive analytics is not yet well defined or established. The Centre for Humanitarian Data plays a key global convening role, and sectoral and geographical specialists are beginning to emerge.
- The use of predictive analytics by humanitarian actors is still an emerging practice characterised by pilot projects and early-stage innovations, which require further development and validation.
- Open data is a significant enabler of predictive analytics in humanitarian action.

Risks and downsides:

- Feeding machine learning with historic data runs the risk of reproducing past errors, prejudices and inequalities.
- Feeding machine learning with social media data runs the risk of amplifying the voices and concerns of the relatively privileged at the expense of the most vulnerable and marginalised.
- Automating algorithmic processes is dehumanising and potentially in conflict with humanitarian commitments to human-centred and participatory processes.
- The need for computing power and data science expertise makes it difficult for small and local actors to lead on predictive analytics – potentially creating new dependencies.
- Together these risks may unintentionally lead to a form of digital humanitarianism that reflects, reproduces and amplifies patterns of historic inequality along intersecting lines including gender, race and class.

Limitations:

- This preliminary mapping and analysis is based on a rapid 14-day desk review of secondary sources, many of which are authored by innovators themselves.
- It was not always possible to verify the claims made, to clarify whether initiatives are still on-going, or to find sufficient detail to answer research questions in detail.
- Additional initiatives and literature continued to come to light even after the cut-off date showing scope for additional mapping.

Recommendations:

- Governments, humanitarian agencies, funders and private companies should publish more open data in order to further extend the potential for predictive analytics.

- Humanitarian agencies should apply the precautionary principle in data collection, data safeguarding and responsible data to protect vulnerable populations from harm.
- To align practice with humanitarian principles and commitments, predictive analytics actors need to include affected populations in all aspects of the design and project cycle.
- Funding of predictive analysis should be tied to risk assessment, risk mitigation and knowledge sharing on the ethics and downside-risks of predictive analytics.
- Funders should support the emerging ecosystem to develop geographical or thematic specialisms, convene knowledge-sharing events and produce ethical guidelines for practice.
- Further research is necessary to build on this preliminary mapping and analysis in this crucial and rapidly developing area of humanitarian action.
- Primary research interviews with humanitarian agencies and key informants would make it possible to validate claims and establish the current status and future plans of initiatives
- A small number of case studies would improve depth of understanding about approaches being used and proposed pathways to scale.
- Focus groups or a workshop would surface agency experience of risks and barriers not shared in publicly accessible documents and enable lesson learning.

2. Introduction

The purpose of the report is to provide a preliminary mapping of the breadth of predictive analytics initiatives being applied to humanitarian action. This initial scoping study is based on a 14-day desk study of secondary sources, many of which were produced by the initiatives themselves and tended to focus on the positive potential of their ambitions rather than on the limitations or challenges. Reviewing so many initiatives in such a short period of time meant that it was not always possible to know whether the initiatives were still on-going or whether they succeeded in realising their early ambitions. A second round of primary research would be an effective way to verify and deepen these preliminary findings.

This brief introduction is followed by an overview of predictive analytics in wider society as context for its application in humanitarian aid in Section 3. Readers familiar with the data collection techniques, machine learning and statistical modelling technologies that underpin predictive analytics may wish to skip this section. In Section 4 we summarise the existing literature on predictive analytics in the humanitarian sector. Section 5 is where we begin to present findings from our review of current applications of predictive analytics in humanitarian practice. The review is based on desk-based research reviewing secondary sources: existing academic literature, grey literature, agency reports and humanitarian websites. In Section 6 we explain how predictive analytics is being applied in humanitarian practice by presenting new information on the range of data sources, data collection methods, and data modelling techniques being used to predict humanitarian emergencies. Section 7 is where we provide a typology of uses of predictive analytics to calculate where and when humanitarian emergencies will occur, who will be the most affected populations, and what the situation will look like. In Section 8 we outline future plans and direction for humanitarian predictive analytics. In the final sections, we review the risks and downsides of predictive analytics in humanitarian practice and make some tentative conclusions and recommendations.

3. Predictive analytics

Predictive analytics involves the recognition of patterns in historic data to calculate the likelihood of future events. Recommendation engines in Netflix, YouTube and Amazon use predictive analytics to recognise patterns in your previous online activity (and that of people with patterns similar to yours) to statistically calculate the probability of which film, video or book you are most likely to want next. Cambridge Analytica and other political marketing consultancies take big data from Facebook and the electoral register and use machine learning to build behavioural profiles of every citizen to predict their voting preferences and micro-target them with political influencing messages. In theory, the more data they have on each individual the more accurate their predictive analytics. Siegel (2016: 15) defines predictive analytics as “technology that learns from experience [historic data] to predict the future behaviour of individuals in order to drive better decisions”. The three main components of predictive analytics are big data, machine learning and statistical modelling.

Big data is often crudely defined as data sets that are too big to be analysed in a standard spreadsheet or too big to fit on a personal computer hard drive. Big data can consist of both structured and unstructured data. Structured data is data that is quantitative in nature and fits neatly into the rows and columns of a spreadsheet such as government statistical records or budgetary information. Unstructured data might include the text of hundreds of different documents, video and photos scraped from social media, Global Positioning System (GPS) mobile phone traces, satellite images and facial recognition images.

Machine learning is the most often used tool in predictive analytics. It is a type of artificial intelligence which is used to find patterns in big data and uses them to calculate the probability of future events. Machine learning takes any explanatory variables that are found to be highly correlated with a particular past outcome and uses them to produce predicted future variables. A statistical model is used to assign a probability ‘score’ to each possibility. This can then be used to predict anything from voting patterns, commodity prices, migration flows or flood trajectories.

Statistical modelling: The statistical analysis can use a single data set and a single model or combine multiple data and multiple scenarios in ‘ensemble models’. Ensemble modelling is the combination of multiple statistical models to improve predictability. Predictive analytics often uses hundreds or thousands of predictive models to analyse the probability of a range of possible future scenarios (Siegel, 2016). The increased availability of computing power, big data, and machine learning makes possible the automation of multiple statistical models at a fraction of the time and cost of traditional data modelling. The predictions generated by statistical modelling can be provided to human decision-makers to inform their deliberations (as with film recommendations that Netflix provides to support your choice of viewing) or the prediction can be used to drive an automated algorithmic decision-making process (as when YouTube auto-plays its video choice for you). Automated analytics is an emerging field where decisions based on predictive analytics are algorithmically determined and implemented entirely automatically (Davenport, 2015; Castellucia & Le Métayer, 2019).

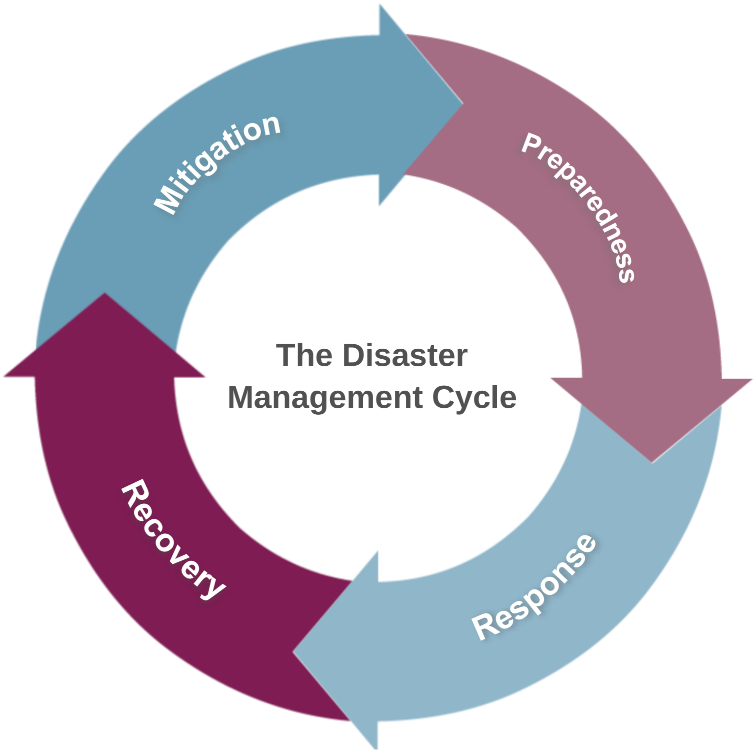
Predictive analytics has limits, comes with risks and raises ethical issues. Predicting that something will happen to a specific individual, community, or geography with 100% accuracy is impossible. It is not clear where legal liability resides if predictive analytics leads to injury or death. Predictive analytics is used in ways that are ethically unsound, for example, to nudge individuals towards thoughts, behaviours and voting preferences without their consent or transparency (as per Cambridge Analytica). There is also a growing research literature documenting evidence that the use of historic data in machine learning and algorithmic decision-making often reflects, reproduces and amplifies historical patterns of gender, race and class (dis)advantage and inequality (Benjamin, 2019; Criado Perez, 2019; Eubanks, 2018; Hernandez & Roberts, 2018; Noble, 2018; O’Neil, 2017). The response of technologists to this (conscious or unconscious) bias in data and the politics in algorithms is often to try to manufacture a technological fix of the data or the algorithm rather than to address the social problem itself or its root causes. It has also been argued that the use of algorithms, artificial intelligence and automated decision-making is dehumanising by definition in that it replaces the scope of human agency, deliberation and dialogue (Roberts & Faith, in press). Bastani and Kim (2018) are among the many scholars who have argued that it is important to keep domain experts engaged in an iterative human process of predictive analysis (the so-called 'humans-in-the-loop' argument).

4. Predictive analytics in humanitarian aid

Humanitarian actors and the humanitarian sector are often criticised for being slow to act and operationally inefficient (Swaminathan, 2018). In response to these challenges, the humanitarian sector has sought to shift from being responsive to disasters and crises to being more anticipatory. This has involved the increased use of early warning and forecasting systems to strengthen disaster prevention, preparedness and mitigation. Predictive analytics holds the potential to extend these proactive capabilities before and during disasters (Akter & Wamba 2019; Swaminathan, 2018).

The disaster management lifecycle consists of four stages: mitigation, preparedness, response, and recovery (Haigh, n.d.). The mitigation stage is designed to decrease the chances of a disaster happening or its potential impact on vulnerable populations and places.

Figure 1. The four phases of the disaster management cycle



Source: Adapted from Haigh, n.d.[1]

[1] Adapted from "Disaster Management Lifecycle", by R. Haigh, n.d. http://www.science.earthjay.com/instruction/HSU/2017_spring/GEOL_308/lectures/lecture_01/GEOL_308_suppl_reading_02_Introduction_to_Disaster_Management_Lifecycle.pdf © University of Salford. Reproduced under licence CC BY-NC-SA 2.5.

The preparedness phase aims to improve readiness for future disasters and includes funds allocation and prepositioning of assets. The response phase occurs after the disaster has hit and humanitarian actors are active on the ground. The recovery phase includes actions seeking to bring about long-term stabilisation after the disaster (Akter & Wamba, 2019). In a systematic review of 76 scholarly articles on big data in disaster management, Akter and Wamba (2019) found that 37% of all articles focused on mitigation, 29% on response, 23% on preparedness and only 3% focused on recovery.

The existing literature highlights the potential for big data to help predict and prevent disasters but that there is a lack of real-world case studies because the use of big data in the humanitarian sector is relatively new. Watson et al. (2017: 17–18) found that although “studies demonstrate that crisis data have the potential to positively impact preparedness, there has been little empirical research relating to the actual use of crisis data for preparedness activities”. A workshop hosted by the Centre for Humanitarian Data in April 2019 also found that “Most of the models that were shared by organizations are in a pilot phase and still need further validation and feedback before they can be used to create a trusted signal for the [humanitarian] sector to respond to” and suggested that predictive analytics models will need to be used alongside existing forecasting techniques until the evidence base is built (Centre for Humanitarian Data, 2019a: 2). The workshop also highlighted the need for case studies and documentation of humanitarian predictive analytic projects.

There are signs that this is changing. In January (of 2020), the International Federation of Red Cross and Red Crescent Societies (IFRC) made its first use of its ‘Early Action Funding Mechanism’ tool to provide cash to vulnerable farmers predicted to loose livestock in a particularly harsh winter (IFRC, 2020). Although still largely in their pilot phases, several other humanitarian predictive analytics projects have been included in grey literature and/or media coverage, including the World Bank’s Famine Action Mechanism, which aims to predict famines before they happen and trigger funding based on predictions to facilitate earlier responses and possibly prevent crises (OCHA, 2019a); Save the Children’s forced displacement prediction tool which aims to provide actionable predictions about how a situation of forced displacement is likely to evolve over time (Morgan & Kaplan, 2018); and the UNHCR’s Jeston project which can predict the displacement of people in Somalia at least a month in advance (UNHCR, 2019).

Although the use of predictive analytics is now widespread in the private and public sector in many developed countries, in the development and humanitarian aid sector the use of big data, machine learning and artificial intelligence are still at an exploratory stage (Paul, Jolley & Anthony, 2018). Although the use of predictive analytics by humanitarian actors is still in its infancy, attempts by humanitarian actors to apply predictive analytics are not completely new. One early proof of concept dates back to 2010 during the aftermath of the Haiti Earthquake, where call records of 1.9 million Haitians were analysed between 1.5 months prior to and almost one year following the earthquake. Results showed that the movement of Haitians within the country could be predicted during the first three months of the disaster (Lu, Bengtsson & Holme, 2012).

Moreover, predictive analytics are not the humanitarian sector’s first attempt at predicting disasters. The sector has long made use of forecasting early warning methodologies. However, as is the case in other sectors, the promise of predictive analytics in its new digital form (e.g. the combination of big data, machine learning and statistical modelling) is to go much further and get there faster than traditional forecasting and early warning methodologies. The following excerpt from a U.S. Agency for International Development (USAID) commissioned report on machine learning captures this well:

> Of course, not all early-warning systems rely on machine learning. It is common for people to analyse geospatial, economic, or health data and make predictions about what might happen. One major difference is that human analysts tend to make predictions based on a small number of strong signals, such as anticipating a famine if rainfall is low and food prices are high. In contrast, machine learning methods excel at combining a large number of weak signals, each of which might have escaped human notice. This gives machine learning-based early warning systems the potential to find the 'needle in a haystack' and spot emerging problems more quickly than traditional methods (Paul et al., 2018: 19).

The literature reviewed found a series of claims for the relevance of predictive analytics in relation to the different phases of the humanitarian lifecycle: in disaster mitigation, preparedness, response and recovery. These claims are summarised below.

Mitigation: Predictive analytics can inform mitigation phase strategies that seek to prevent disasters and/or crises from happening or limit their impact once they happen (Akter & Wamba, 2019). Predictive analytics can be used to calculate vulnerability to natural hazards and pinpoint which households, communities, and infrastructure humanitarian actors should prioritise (Letouze, Sangokoya & Ricard, 2017). The idea of using vulnerability as a predictor is not new. "vulnerability has a predictive aspect: it should be possible—on the basis of the characteristics of a group of people who are exposed to a particular hazard—to identify their capacity for resilience" (Cannon, 2008: 10). Locating vulnerable people and geographies vulnerable to natural hazards can be crucial as it is vulnerability to hazards that lead to disasters rather than the hazards themselves (Cannon, 2008). Addressing these vulnerabilities, therefore, has the potential to prevent or limit disasters.

Preparedness: In the planning phase, predictive analytics can provide actionable early warning to authorities, citizens and humanitarian actors about imminent threats (Akter & Wamba, 2019). What is considered early will vary for different disasters and crises. For example, it may be possible to predict a famine months ahead of time, but it may only be possible to predict which areas are under threat of flooding due to a hurricane a week in advance or an earthquake just minutes before it happens (Watson et al., 2017). Hala Systems, one of the initiatives uncovered during our mapping exercise predicts which areas of Syria will be bombed by military planes 5 to 10 minutes before the bombs land, sending text and instant messages to those in the affected locations (Hala Systems Inc., 2019). This very narrow time window provides citizens with just enough time to take shelter.

Response: During the response phase, predictive analytics can help provide situational awareness. The use of predictive analytics at the response stage is strongly related to 'now-casting' which refers to making real-time inferences about what will happen in the short term based on data (Letouze et al., 2017)."In the short term, the information gained from social media and other aerial imagery has the potential to inform those managing a crisis, who and where vulnerabilities might lie as a crisis develops. This could include 'trend analysis' and 'predicting which populations are vulnerable' to health [and other] risks, abuse or other additional effects" (Watson et al., 2017: 19–20). Predictive models based on call records and GPS data has been used to predict where people are most likely to flee or relocate (Lu et al., 2012). Predictive analytics can also provide early assessments of damages and losses (e.g. by analysing and classifying satellite imagery of the roofs of people's homes) providing humanitarian actors with much needed data to guide rapid response (Letouze et al., 2017).

Recovery: Relatively little work has been done to develop thinking on how predictive analytics could be used during recovery efforts. Echoing findings from Akter and Wamba (2019) our literature review found that research on predictive analytics in the humanitarian sector lacks studies on the use of predictive analytics in the recovery phase of humanitarian action. In our mapping of the 49 cases of humanitarian predictive analytic initiatives, we found even coverage in the mitigation, preparedness and response phases of the humanitarian lifecycle. However, we did not find any initiatives focusing on the recovery phase.

Like in other sectors, humanitarian predictive analytics is operationalised using a combination of big data, machine learning, and statistical modelling. Humanitarian emergencies leave behind a substantial trail of big data. The data from one disaster can be reused to inform mitigation, preparedness, and responses to future disasters (Watson et al., 2017). However, the humanitarian sector does not need to rely exclusively on the data that it produces itself. A large variety of external big data sources can be used to feed machine learning and statistical models to produce predictive analytics to guide humanitarian action. Data sources can include household surveys, mobile phone metadata, satellite imagery, social media, financial transaction data, crowdsourced crisis data, weather data and data held by government and private companies.

5. Mapping humanitarian predictive analytics

This section begins with the presentation of findings from our own research. It begins with an overview of the actors in the ecosystem of humanitarian predictive analysis. The mapping was limited by the lack of detail in the secondary data available online about many of the initiatives. In many cases, the only available data was written by project actors themselves. Much of the data was incomplete and insufficient to fully answer our research questions. A second round of primary research with innovators and informant interviews would strengthen the depth and validity of the data and enable research questions to be answered more thoroughly.

Key organisations

A key convening and coordinating role at the global level is currently played by the Centre for Humanitarian Data hosted at the United Nations (UN) Office for the Coordination of Humanitarian Affairs (OCHA). The Centre surveyed thousands of humanitarian professionals in 2019 and found that big data, predictive analytics, and statistical modelling were the top three areas that humanitarian workers wanted to learn more about (Centre for Humanitarian Data, 2019b). This led the Centre to begin a work-stream dedicated to predictive analytics. Its work on predictive analytics has three main goals: to develop new predictive models, provide quality assurance through peer review,[2] and create a community of practice where knowledge can be shared, and partnerships and collaboration can be built (OCHA, 2019b). The Centre convenes events, provides training and helps set standards (OCHA, 2019b). The centre has crowdsourced a list of predictive analytics projects by humanitarian organisations.[3]

While the Centre for Humanitarian Data plays a global role across all humanitarian sectors, there are a number of other initiatives emerging with sectoral or geographic specialisations. The Precision Public Health initiative is a $100 million fund seeking to scale and accelerate existing predictive analytics solutions related to health including in the area of disease outbreaks. The fund is supported by the Rockefeller Foundation in partnership with the United Nations Children's Fund (UNICEF), the World Health Organization, ministries of health, global health agencies, and

[2] https://centre.humdata.org/wp-content/uploads/2019/09/predictiveAnalytics_peerReview_updated.pdf

[3] https://centre.humdata.org/catalogue-for-predictive-models-in-the-humanitarian-sector/

technology companies. The fund has two main goals. Firstly, to invest in initiatives that use “predictive analytics to prevent rather than respond to health threats and enable them to be more targeted, effective and efficient” Secondly to “invest in leveraging big data on the social determinants of health—to identify the populations at greatest risk in order to facilitate the delivery of health interventions” (Rockefeller Foundation, 2019: 2). The initiative is also seeking to create an enabling environment for predictive health analytics by helping to address gaps in data availability and quality, data science talent, and policies regarding the responsible use of data.

UN Global Pulse focuses on the use of big data in development, humanitarian aid, and peacebuilding sectors.[4] UN Global Pulse is a partner to several of the humanitarian predictive analytics projects identified in this study. Data-Pop Alliance is a collaboration between Harvard Humanitarian, MIT, the Overseas Development Institute (ODI) and the Flowminder Foundation which seeks to “bring together researchers, experts, practitioners, and activists”.[5] The Big Data for Development (BD4D) network is “a Southern-led partnership [partnership between the African Institute for Mathematics Sciences (AIMS), Local Development Research Institute (LDRI), Centro de Pensamiento Estrategico Internacional (CEPEI), LIRNEasia, and the Centre for Internet and Society (CIS) founded with the objective of developing policy-relevant research on big data for development that is conceptualized and implemented by Southern organisations”.[6] A final relevant initiative is the private sector ‘AI for Impact’ programme of GSMA which seeks to scale the use of mobile big data analytics to tackle challenges related to the Sustainable Development Goals (SDGs).[7]

We used the Centre for Humanitarian Data’s existing crowdsourced catalogue of 24 humanitarian predictive analytic projects as a point of departure for our research (although some projects uploaded to the Centre for Humanitarian Data’s site did not use predictive analytics in humanitarian work or did not include enough information to enable inclusion). Twenty-one of the 24 initiatives in the catalogue were included in our mapping, contributing 43% of the initiatives that we mapped and analysed. Two initiatives were removed because they did not make predictions based on historical data and one initiative was removed due to a lack of information. The addition of 28 new initiatives was produced through desk-based research of secondary sources, internet search and personal referrals. Additional projects continued to emerge after our cut-off date (May 1, 2020) which we were therefore unable to review or include. Our list is partial and preliminary. The field is growing and scope for further mapping exists.

Based on the initiatives that we mapped, the following sections summarise what kind of organisations are active in humanitarian predictive analytics, which countries they work in, and what sectors of humanitarian work they are focused on.

Organisation type

Our research identified 49 initiatives that use predictive analytics in humanitarian work. The humanitarian predictive analytics space is diverse in terms of types of organisation leading projects. Humanitarian Agencies lead 13 of the initiatives (27%) of which nine were partnerships

[4] https://www.unglobalpulse.org/

[5] https://datapopalliance.org/about/vision-and-members-2/

[6] http://bd4d.net/

[7] https://www.gsma.com/betterfuture/aiforimpact

with other organisations. Multilateral organisations including UNICEF and the World Bank account for eight of the initiatives (16%) of which seven were partnerships with other actors. Ten initiatives (20%) were led by private companies, including Microsoft and Google and small start-ups, of which only four were carried out in partnership with other actors.[8] Nine initiatives (18%) were led by academic institutions. Of the academic-led initiatives, five consisted of partnerships. Three initiatives (6%) were led by international non-governmental organisations (INGO) that do not primarily focus on humanitarian action (e.g. Oxfam), of which one was a partnership with other actors. Two initiatives were led by Intergovernmental Organisations, namely the European Union and African Union, all of which were partnerships with other actors. Only two initiatives (4%) were led by bilateral donors notably DFID's Cholera Prevention programme in Yemen (in partnership with OCHA, UK Meteorological Office, U.S National Aeronautics and Space Administration (NASA), University of Maryland and University of West Virginia) and USAID's Famine Early Warning Systems Network (FEWSNET) (in partnership with NASA, the National Oceanic and Atmospheric Administration [NOAA], the U.S. Department of Agriculture [USDA], and U.S. Geological Survey [USGS]). Only one initiative was led by a national or local government body, namely the Mongolia dzud model which is a partnership between the National Agency for Meteorology and Environmental Monitoring and the IFRC. Lastly, one proof-of-concept initiative (2%) was led and carried out by an individual person without any clear links to an organisation at the time of publishing.

Table 1 provides an overview of the number and percentage of initiatives led by each type of organisation and their use of partnerships.

Table 1. Humanitarian predictive analytics initiatives by organisation type

Lead organisation type	Number of initiatives	% of total	Of which are partnerships	% that were partnerships
Humanitarian organisation	13	27%	9	69%
Private	10	20%	4	40%
Academic institutions	9	18%	5	56%
Multilateral organisation	8	16%	7	88%
INGO	3	6%	1	33%
Bilateral donors	2	4%	2	100%

[8] We did not consider funders or clients as partners.

Intergovernmental organisation	2	4%	2	100%
Local government organisation	1	2%	1	100%
Individual person	1	2%	0	0%
Total	**49**		**31**	**63%**

Source: Authors' own.

Although partnerships were common in the initiatives that were mapped, we were unable to locate information about the types of partnerships, their formats and relative utility. It was also noticeable that partnerships did not generally include a role for populations affected by humanitarian emergencies. Populations affected by conflict and disaster have the right to inclusion in projects intended to benefit them[9] and they possess invaluable contextual and cultural knowledge. Humanitarian innovation too-often fails to build on indigenous knowledge, local innovation and existing capacity. To be optimally effective, humanitarian innovation needs to be in local languages, reflect local culture, be promoted by local champions, and adopted and sustained by local organisations. The International Committee of the Red Cross (ICRC) in their contribution to the UN Secretary-General's High-Level Panel on Digital Cooperation warned that it is critically important that affected populations are "firmly at the centre of any initiatives in order to ensure the humanitarian response *do no harm* [emphasis added] in their application."[10]

Sectoral application

Our mapping found that predictive analytics is being applied across a wide variety of humanitarian sectors. There were nine initiatives each in the areas of *disease outbreaks* and *displacement and migration*. Seven initiatives focused on *conflicts*. Six initiatives were concerned with *disaster risk reduction*. Four focused on *food security*. *Floods; human resources*; *logistics; water, sanitation and hygiene (WASH)*; *nutrition*; and *disaster response* accounted for two projects each. *Fundraising* and *livelihoods* each had one project as well. Table 2 summarises these findings. In addition to applying predictive analytics to humanitarian operations, our study found that 10% of initiatives used predictive analytics in back-office activities (human resources, fundraising and logistics).

[9] http://www.inclusioncharter.org/

[10] https://www.digitalcooperation.org/wp-content/uploads/2019/02/Charlotte_Lindsey_Curtet_CFC-ICRC-1.pdf

Table 2. Predictive analytics initiatives by humanitarian sector

Sector	Initiatives
Disease outbreaks	9
Displacement and migration	9
Conflict	7
Disaster risk reduction	6
Food security	4
Floods	2
Human resources	2
Logistics	2
WASH	2
Disaster response	2
Nutrition	2
Fundraising	1
Livelihoods	1
Total	**49**

Source: Authors' own.

Geographical application

The initiatives identified were geographically dispersed, but the majority identified in this study were focused in Africa and the Arab world with significant numbers in Asia and Latin America. We were unable to verify the extent to which this is explained by overall funding patterns or other factors.

Four Initiatives (8%) focused primarily on the sub-national level including three initiatives with a focus on a single state or city within a country namely in Bihar (India), Cali (Colombia), and Chicago (USA) and one initiative with a focus on a single refugee camp, namely the Za'atari camp in Jordan which is home to 80,000 Syrian refugees. There were 13 national-level initiatives (17%) including initiatives in Yemen (3), Kenya (2), Malawi (1), Syria (2), Madagascar (1), Mongolia (1), Pakistan (1), Philippines (1), and Somalia (1).

Eight initiatives (16%) covered a few countries (2–5) within a region. Of these, five were in Africa and three in Asia. Five initiatives (10%) sought to make predictions in many countries (e.g. 10 countries within a region). These were evenly spread across the Arab world, Africa, Asia, and Latin America. Two initiatives (4%) focused on between 2–5 countries in more than one region. One of these initiatives consisted of countries across Africa and Asia while the other consisted of countries across Asia and Latin America.

Nine initiatives (19%) have global scope, while a further three (6%) sought to make predictions in many countries geographically spread across the world including one initiative that aims to make predictions for all lower- and middle-income countries (LMICs). A further three initiatives (6%) operated at an organisational level including Médecins Sans Frontières (MSF)'s People Analytics human resources project and Direct Relief's logistics predictive analytics project in partnership with IBM.

Table 3. Humanitarian predictive analytics initiatives by geographical focus

Geographical focus	Number of initiatives	% of total
National (17)		
Sub-national location	4	8%
National	13	27%
Regional (17)		
2 to 5 countries in a region	8	16%
More than 5 countries in a region	5	10%
2 to 5 countries in 2 regions (e.g. Asia and Africa)	2	4%
Geographical intergovernmental areas (e.g. European Union and African Union)	2	4%
Global (15)		
Many countries around the world but not fully global (e.g. 20 countries)	3	6%

Fully global	9	18%
Global at the organisational level	3	6%
Total	49	

Source: Authors' own.

6. Approaches to predictive analytics

This section presents findings on the range of technologies and technical approaches to predictive analytics currently being used by humanitarian actors. It covers data sources, data collection methods and approaches to modelling prediction. The findings in this section are limited by the fact that the publicly available materials often did not detail the approach being used. Some initiatives do not share any details of their approach, others provide only a brief explanation insufficient for confident categorisation, whilst others document their methodology in detailed research papers and journal articles. Due to these limitations in the data, we cannot say precisely how many of the mapped organisations use a specific methodology. Primary research would be necessary to increase the quality of these findings. However, we are able to report the number of initiatives that did detail their technical approach.

The examples of humanitarian predictive analytics that we identified are being used to complement, rather than to replace, traditional humanitarian analysis and forecasting. Most of the approaches rely on historical data which is collected using traditional survey and data collection methods. The primary data sources are humanitarian agencies' own in-house data sets such as household surveys and open government data. As examples below illustrate, we also found instances of predictive analytics being used to complement existing early-warning and forecasting methods as a means of making predictions about some of their data blind spots i.e. areas where household surveys are not practically possible. For example, the World Food Programme (WFP)'s Hunger Map Live already makes forecasts about the state of food security around the world but is often unable to conduct household surveys in conflict zones. To overcome this WFP uses predictive analytics to address these blind spots on the map based on what is happening nearby and in areas with similar characteristics.[11]

Data sources

This study found that humanitarian predictive analytics projects are making use of a diverse and expanding range of data sources and types. In some cases, the historical data consisted of data that the organisation was already collecting in-house. As should be expected, this was especially prevalent in humanitarian and multi-lateral organisations which have a long history of collecting historical data. For example, WFP is constantly collecting food security data through household surveys which in recent years have been digitised and automated creating a large pool of big data. WFP has partnered with the Chinese e-commerce and logistics company Alibaba to analyse WFP's Computer Assisted Telephone Interview (CATI) data to make real-time hunger

[11] https://hungermap.wfp.org/

predictions on Hunger Map Live.[12] Similarly, IBM is using the Danish Refugee Council's pre-existing interview data to help predict patterns of population movement in their Mixed Migration Foresight Project (Danish Refugee Council & IBM, 2018). Other traditional data sources included case studies, expert interviews, and socio-economic household survey data.

Open data is used heavily in humanitarian predictive analytics initiatives. Although we found only seven instances where open data was mentioned explicitly, there were many other situations where open data appears to be implicit. Data sets made open and freely available in machine-readable formats by the World Bank and UN Agencies were mentioned by several initiatives. Open data was shown to be a significant enabler of predictive analytics in humanitarian action so increasing the availability of open data extends the potential of predictive analytics in humanitarian aid. Further research could usefully establish the precise extent and utility of open data in humanitarian predictive analytics.

Satellite imagery was one of the most popular data sources found in at least nine (19%) of the initiatives. This popularity is explained partly by the increasing availability of free and open access satellite imagery.[13] This study found satellite imagery being used to make humanitarian predictions across many sectors. UKAID and partners have successfully used satellite imagery to help predict which areas are at risk of a cholera outbreak in Yemen 6 days in advance. Google feeds satellite imagery to its 'inundation model' to help predict which areas are at risk of flooding (Nevo, 2019). WFP's 'Platform for Real-time Impact and Situation Monitoring' (PRISM) combines satellite imagery with data gathered from remote sensors to predict climate hazard risks (World Food Programme, 2020). The Internal Displacement Monitoring Centre (IDMC), United Nations Office for Disaster Risk Reduction,and ETH Zurich are using satellite imagery to help predict displacement and damages after natural disasters by mapping houses that show sign of roof and house damage (Milano, 2017). The African Risk Capacity, "an advanced satellite weather surveillance and software – developed by [WFP] – to estimate and trigger readily available funds to African countries hit by severe weather events" (African Risk Capacity, n.d.).

Social media data was used in at least four initiatives. The Crisis Computing Team at the Qatar Computing Research Institute (and partners) apply natural language processing and computer vision techniques to social media "to crisis information communicated via social media to gain situational awareness and actionable information for the humanitarian response" (QCRI, 2017). The United Nations High Commissioner for Refugees (UNHCR) Innovation Service, UN Global Pulse, and Crimson Hexagon looked into the feasibility of using Twitter data to analyse how people feel and think—known as a sentiment analysis—about the arrival of refugees in Europe to counter misinformation and to understand how refugees themselves felt about their new living conditions (UNHCR, 2019; UN Global Pulse, 2017). The International Red Cross and Red Crescent Movement has used Facebook population data to help map exposure to risks for its forecast-based financing programme (International Red Cross and Red Crescent Movement, 2019). Humanity Data Systems—a commercial start-up—combines social media data with many different data sources to help simulate active conflict environments and determine how best to engage in the context (Humanity Data Systems, 2019). The use of social media data comes with known risks. Social media users are disproportionately urban, middle class, and literate so there is a danger that listening to their voices drowns out the voices of more disadvantaged populations (Roberts & Marthais, 2017). Privacy International has warned that humanitarian

[12] https://hungermap.wfp.org/

[13] https://eos.com/blog/7-top-free-satellite-imagery-sources-in-2019/

organisations using social media metadata risk that use leading to surveillance, false identification, and mistargeting.[14]

Not covered explicitly in the majority of initiatives' documentation was the crucial issue of data safeguarding. Holding and processing data on vulnerable populations and their locations creates new risks and responsibilities for humanitarian actors. In a recent workshop in London on artificial intelligence in humanitarian aid, senior staff confessed that their data-management processes were sub-optimal and that the potential for data breach was a great concern (Roberts & Faith, in press). Big data comes with bigger risks and tackling the issues of responsible data (O'Donnell, 2015) and data safeguarding practices needs a higher profile in humanitarian predictive analytics.

Data collection methods

Crowdsourcing was used in at least three initiatives as a data collection method to feed predictive model. The company Premise Data and the local government of Cali, Colombia are relying on crowdsourced citizen data to predict potential Zika and other mosquito-transmitted disease hotspots (Novak, 2018). Similarly, the Cholera Artificial Learning Model (CALM) initiative have developed a chatbot to obtain rapid WASH data from places affected by cholera outbreaks (Badkundri et al., 2019). The Simon-Skjodt Center and Dartmouth College have used "crowd forecasting tools—including a public opinion pool and annual comparison survey—[to] leverage the 'wisdom-of-the-crowd' and provide real-time assessments" among with many other sources of data to predict the risk of mass killing and genocide across the world in their early warning project.[15]

Remote sensing was used to generate data in at least five humanitarian predictive analytics models. UNICEF is applying a 'smart city approach' to a refugee camp in Jordan's freshwater movement system through the application of sensors and laser meters to help predict which water tanks need to be refilled and when (Kaplan, 2017). Hala Systems Inc.—a start-up—uses sensors that track noise made from aeroplanes to determine which areas are likely to be bombed in the next 5-10 minutes and send alerts to citizens (Hala Systems Inc., 2019).

Drone images were the data collection method in at least three humanitarian predictive analytics initiatives in our mapping. UNICEF piloted a project that combined data collected from drones with field observations to predict areas at high risk of malaria in Malawi (Stanton & Jones, 2018). Mercy Corps has looked into the use of drones to help map flood risk in Nepal and Timor-Leste (Sterling, 2017).

Data analysis methods

Mapped initiatives that were explicit about their methodologies all combined historic data and new data sources with new machine learning or modelling technologies. This section details the data analysis methods being used.

Machine learning was referenced as an approach by 35 initiatives (73%) or implied by the use of the more general term artificial intelligence. Although the existing literature suggests that machine learning is a central element in most digital predictive analytics, we found 13

[14] https://privacyinternational.org/news-analysis/2535/do-no-harm-digital-age-privacy-and-security-cannot-be-ignored

[15] https://earlywarningproject.ushmm.org/about

humanitarian initiatives (27%) that claimed to be doing predictive analytics without mentioning of machine learning. It is possible that organisations are using different definitions of predictive analytics. Of the 13 initiatives for which use of machine learning was unclear, six mentioned the use of statistical modelling, and two referred only to 'software'. We would need to interview project actors to overcome these categorical limitations.

Natural language processing is a subfield of artificial intelligence using tools that can process, analyse and respond to data containing human language (e.g. chatbots, Siri and Alexa). Natural language processing was used by at least three of the initiatives mapped. For example, the IDMC uses natural language processing to scrape through 5,000 media articles per day to extract 2,000 displacement facts per day which it uses to feed predictive models (IDMC, n.d.). The Crisis Computing Team at the Qatar Research Institute (and partners) apply natural language processing and computer vision techniques (algorithms that analyse digital images and videos) "to crisis information communicated via social media to gain situational awareness and actionable information for the humanitarian response" (Crisis Computing team at Qatar Computing Research Institute, 2017).

Facial recognition software refers to technology that can identify individuals by analysing their facial features. One initiative by the start-up company Kimetrica is using predictive analytics and facial recognition software to predict which children are likely to be malnourished.[16]

Computer simulations that predict the most likely outcomes for multiple scenarios are being used by at least six of the projects identified. For example, Predictify.Me was using computer simulations to predict which schools were at risk of a terror attack and how well the schools were prepared (Army Technology, 2015).[17] Uppsala University's Violence Early Warning System (ViEWS) uses dynamic simulation techniques to "predict armed conflict involving states and rebel groups, armed conflict between non-state actors, and violence against civilians … at [the] national level, subnational level and actor level" (Hegre et al., 2019: 156).

Scenario building is similar to simulations but is done manually rather than generated automatically by an algorithm or computer software. At least three of the projects identified were analysing alternative potential scenarios that may warrant different responses. USAID's FEWSNET utilises if/then scenario development to develop a 'most-likely scenario' of food security and what shocks or events may change the scenario in which ways (Famine Early Warning Systems Network, n.d.).

7. What is being predicted?

This section presents our findings on what humanitarian problems predictive analytics is being used to address. It provides many concrete examples of what is being predicted and why.

Our original plan was to categorise humanitarian predictive analytics initiatives by the phase of the humanitarian cycle that most strongly aligned with each initiative. The mapping exercise highlighted a major issue with this approach: most initiatives claimed to address two or three of the phases. We were unable to categorise every initiative due to information gaps. Those that provided sufficient data were spread relatively equally across mitigation/prevention (21),

[16] https://kimetrica.com/our-projects/

[17] It is unclear whether this initiative is still operational. The company itself seems to have closed its doors and its website no longer works.

preparedness (21), and response (14). We did not find any initiatives focusing on the recovery phase. Another 10% of initiatives addressed issues outside of the humanitarian management lifecycle categories in areas such as fundraising and human resource management. Twenty-three projects used language that suggested predictive analytics would improve efficiency by saving time and money for them and partners.

Instead of using phases of the humanitarian lifecycle, it proved more useful to categorise initiatives according to the issue that they were predicting. We found that humanitarian predictive analytics initiatives mainly focus on predicting (a) the most affected populations, (b) the characteristics of the humanitarian situation on the ground, (c) the location of humanitarian events and (d) the timing of those events (or who, what, where, when). This typology was more revealing of how predictive analytics is serving humanitarian objectives in practice. Although this typology has not previously been used for predictive analytics Swaminathan (2018: 1697–1698) suggests that big data can help answer compelling who, what where and when questions in many situations.

Our analysis of initiatives illustrates that humanitarian predictive analytics projects are currently focused on answering four primary questions: (i) **WHO** will be most at risk in the advent of a disaster? (ii) **WHAT** will the situation be like? (iii) **WHERE** will events requiring humanitarian action likely unfold? (iv) **WHEN** will humanitarian deployment be needed? Figure 2 illustrates this typology including further related questions.

Figure 2. Examples of the who, what, where and when of humanitarian predictive analytics

WHO	WHAT	WHERE	WHEN
• How many people will require humanitarian assistance? • How many people will be displaced? • Who will commit an attack? • Who will likely be left out? • Who will quit? • WHo will support your cause?	• What will the situation be like on the ground? • What will be the likely impact of emerging external factors (e.g. currency devaluation, recessions, droughts, conflict, etc.) on the status quo? • What might the situation look like in data blind-spots? • What will be the likely impact of potential courses of action?	• Where should early funding be released/triggered? • Where will an event requiring humanitarian action take place? • Where will people most at risk be located? • Which areas are at risk of a disease outbreak? • Where will there likely to be a armed conflict? • Which places are likely to be attacked? • Where will displaced people go, where will they come from, and what routes will they take?	• When will the crisis most likely occur or peak? • When will the crisis end/How long will it last?

Source: Authors' own.

Some initiatives targeted more than one question. For example, we came across one initiative that cut across all the questions in our typology. Action Against Hunger's Modelling Early Risk Indicators to Anticipate Malnutrition (MERIAM) project seeks to predict "**who** may be most at-risk of becoming wasted, **when** they are likely to become wasted, and **where** (in what geographical area they reside)” as well as generate “scenarios [the **what**] that demonstrate how stresses and shocks affect this risk at a local level”.[18]

The remainder of this section provides examples of humanitarian predictive analytics initiatives answering the above questions and the humanitarian problems that they address. Where initiatives address more than one question, they were covered under the question that they most strongly aligned with.

Table 4. Number of humanitarian predictive analytics initiatives answering who, what, where, when questions[19]

	At Least
Predict Where	35
Predict Who	20
Predict What	13
Predict When	9

Source: Authors' own

Predicting who

Who will be most affected? This was the second most common question asked in our typology by the initiatives we uncovered, with at least 20 initiatives (41%) aiming to make predictions related to the ‘Who’ portion of our typology. This mainly involved predicting how many people in which demographic groups are likely to be affected as well as who will be most vulnerable, and who is most likely to be left behind. This information is key to enabling humanitarian actors to better target their efforts to aid affected populations by putting in place mitigation, preparedness and response resources. Less expected applications of predictive analytics in this category included the calculation of who among agency supporters is most likely to donate money and who among agency staff is most likely to resign. The remainder of this section details initiatives using predictive analytics to determine the most likely affected populations in humanitarian work.

How many people will be in need? This question was common amongst humanitarian predictive analytics initiatives focusing on natural disasters. OCHA’s Joint Analysis of Disaster Exposure (JADE) is an analysis carried out within 24 hours of a sudden onset emergency which helps humanitarian actors respond by predicting economic and population impacts by identifying the number of people living in affected areas, the number of people of living in the worst affected areas and the number of people who were already vulnerable before the disaster. Getting predictions early after an onset disaster helps humanitarian actors “gauge level of response in early stages and mobilize resources to deliver life-saving aid in a timely manner” (JADE, n.d.). Disease outbreak initiatives are also seeking to answer this question. The CALM initiative has

[18] https://knowledgeagainsthunger.org/research/prevention/meriam-modelling-early-risk-indicators-to-anticipate-malnutrition/

[19] Some initiatives are included under more than one question.

proven effective in forecasting “the exact number (with an error margin of 4.787 cases per 10,000) of cholera cases any given Yemeni governorate will experience for multiple time intervals ranging from 2 weeks to 2 months” (Team Lambert, 2018).

Who will be displaced? This question was commonly asked in predictive analytics programmes focusing on population displacement and migration. The IDMC’s Global Displacement Risk Model seeks to predict the number of people likely to flee damaged homes due to specific natural hazards in a specific location every given year, decade or century. The IDMC hopes these predictions can help “prevent future displacement and reduce its impacts on people and prepare for, and respond to, disaster-related displacement”.[20]

How many displaced people are likely to arrive? Related to the above, answering this question helps humanitarian actors better prepare and plan for influxes of people in specific locations. UNHCR and UN Global Pulse’s Project Jetson uses predictive modelling to better understand “the nexus between displacement, climate/weather anomalies and changes and violent conflict” and thus predict the number of internally displaced people and refugees to and from Somalia.[21]

Who will be left behind? Researchers showed that predictive analytics could be used to determine which children are most likely to drop out of immunisation programmes. Such information could be used to help prevent future outbreaks by targeting at-risk children (Chandir et al., 2018).

Who will attack who? Uppsala University’s ViEWS aims to predict which specific actors will commit an attack. Specifically, the initiative seeks to predict “armed conflict involving states and rebel groups, armed conflict between non-state actors, and violence against civilians … [and] can analyse violence at national level, subnational level and actor level” (Hegre et al., 2019: 156).

Our research also identified the use of predictive analytics by humanitarian actors in their back-office operations including fundraising and human resource activities.

Who will contribute to humanitarian causes? Some humanitarian organisations are now using predictive analytics to mine data on their volunteers to determine under what conditions (e.g. time and place) each volunteer is most likely to offer their help so that they could better target volunteers during the preparation for and response to crises (Fox, 2019). NGOs and humanitarian organisations have also begun to use predictive analytics to sift through their previous fundraising history and donor data to determine which potential funders are most likely to donate for specific causes at specific points in time (BKV, 2013).

Who is likely to resign? Humanitarian organisations are beginning to use predictive analytics to help them better manage their human resources and prepare for future staffing needs and potential employee turnover. MSF has a ‘people analytics’ programme which seeks to identify key staff at risk of leaving allowing MSF to offer potential leavers better offers or promotions in hopes to retain them (MSF, 2018).

[20] https://www.internal-displacement.org/disaster-risk-model

[21] http://jetson.unhcr.org/

Predicting what

Predictions related to the 'What' portion of our typology was the third most common question answered by the initiatives we uncovered. At least 13 initiatives (27%) aiming to make predictions related to the 'What' portion of our typology.

What is the situation on the ground? Along with predicting where, who and when, humanitarian predictive analytics is often focused on providing situational awareness by predicting when supplies will run out as well as what factors are most likely to affect future emergency scenarios. This section contains examples of humanitarian predictive analysis where situational awareness is the prime focus (rather than one among many objectives).

UNICEF Jordan has piloted the use of predictive analytics in urban waste management in the Za'atari refugee camp. The solution allows UNICEF to run more efficiently by predicting the fill rate and status of water tanks to improve humanitarian response and accountability for residents (Kaplan, 2017). Direct Relief is working with General Electric to improve humanitarian response by using data from previous disasters to predict what supplies will be needed where and by whom when a new disaster arises (McKechnie & Axelson, 2019).

USAID's FEWSNET programme helps humanitarian actors create scenarios using if/then analysis. It uses statistical modelling to predict what food insecurity scenarios are likely depending on a range of potential shocks such as currency devaluation, flooding, drought, etc. allowing humanitarian actors to better prepare (FEWSNET, n.d.). The International Crisis Group's CrisisWatch initiative collates up-to-date information about 80 existing and emerging conflicts around the world and provides monthly early warnings of deteriorating situations that may require early action and prevention measures[22] (International Crisis Group, 2016).

What is the situation in data blind-spots? Several initiatives are using predictive analytics to predict what will happen in situations where traditional data collection is impossible. WFP and Alibaba's Hunger Map Live tracks and predicts hunger in near real-time in order to identify areas currently food insecure or predicted to become food insecure. Hunger Map Live uses predictive analytics to predict food security outcomes for blind-spots on its map based on extrapolations from what is happening nearby and in areas with similar characteristics. [23] Brunel University London's Flee Project uses predictive analytics to predict the extent of displacement in places where data collection is incomplete (Buchanunn, 2017). The Drones vs. Mosquitos project by Lancaster Medical School, Liverpool School of Tropical Medicine, Malawi-Liverpool-Wellcome Trust Clinical Research Programme and UNICEF mapped mosquito breeding grounds to predict the situation in other unmapped areas. These predictions could be used in efforts to prevent malaria outbreaks before they occur. Dimagi Inc. and The Arnhold Institute for Global Health's *Integrated Platform to Identify Malaria Data "Cold-Spots"* project seeks to predict the situation in data blind spots where there is a lack of data in the border area between Senegal and Gambia (Gates Foundation, n.d.).

What will be the effect of our action? Predictive scenario building and simulations are increasingly used as part of models to calculate what the likely outcomes will be for any given policy interventions. The Intergovernmental Authority on Development's (IGAD) Conflict Early

[22] It is unclear if predictive analytics tools are used in the production of Crisis Watch's reports. Follow up work is needed to determine if this initiative is in fact a predictive analytics project.

[23] https://hungermap.wfp.org/

Warning and Response Mechanism (CEWARN) programme develops scenario predictions based on input from many partners to inform potential effective preparedness, prevention and response in conflict situations. Their efforts have been "credited with a significant reduction of violent conflict particularly along Kenya-Uganda as well as Ethiopia-Kenya-Somalia borders" (CEWARN, n.d.: 8).[24] Brunel University's Flee project simulates multiple policy options to predict how border closures or camp (re)locations would affect migration (Buchanunn, 2017). Similarly, IDMC's Pastoralist Livelihood and Displacement Simulator allows decision-makers to simulate different humanitarian interventions and predict their ability to withstand droughts, floods, and other effects of climate change, which helps to prevent, mitigate and respond to the impacts of droughts (Ginnetti & Franck, 2014). Private start-up Humanity Data System's Humanitarian Operations Planning Environment (HOPE) simulates specific high-risk conflict scenarios to predict risk factors and help humanitarian actors think through response options of resource deployment (Humanity Data Systems, 2019).

Predicting where

Location prediction is the most common use of predictive analytics in humanitarian aid. At least 35 of the initiatives (71%) used predictive analytics to calculate the most likely location of future humanitarian needs.

Where should early funding release be triggered? The International Red Cross and Red Crescent Movement's Forecast-based financing initiative—now implemented by at least 22 national societies—predefines trigger data points that suggest imminent disasters which if correctly predicted may trigger funding release in time to prevent it from becoming a disaster or mitigate its potential impacts, and to help prepare for and respond to them (International Red Cross and Red Crescent Movement, 2019). Similarly, the World Bank, OCHA, ICRC, Microsoft, Google, and Amazon Web Services have teamed up on the Finance Action Mechanism (FAM) programme which predicts "subnational signal of food insecurity at least 6 months in advance" (Centre for Humanitarian Data, 2020; OCHA, 2018). The predictions from FAM are used to facilitate early action in the form of automatic "pre-agreed, pre-negotiated, pre-arranged" financing and joint response in hopes to increase preparedness and prevent famine (OCHA, 2018). REACH and Shelter Cluster Yemen's 'REACH Flood Susceptibility Model' seeks to help identify areas in Yemen most susceptible to flooding at a 60-metre resolution. It is suggested that the model "can serve to inform humanitarian programming with relation to flood risk [mitigation] and preparedness" (REACH, 2019: 1).[25] The Mongolian National Agency for Meteorology and Environmental Monitoring is able to predict which pastoralists are at risk of losing their livelihoods by ranking household vulnerability and laying it on top of an exposure map and then predicting the outcomes for different groups of people for forecasted natural hazards. Earlier this year, the Mongolian Red Cross society used these predictions to activate its Early Action Protocol (EAP) triggering funding in the form of unrestricted cash assistance and livestock nutrition kits for 1,000 pastoralists deemed vulnerable to an upcoming harsh winter. WFP's Platform for Real-time Impact and Situation Monitoring "assesses the potential risk and forecasts the impact of climate hazards on the most vulnerable communities, in order to design risk reduction [mitigation]

---

[24] It is unclear whether the software used by CEWARN includes predictive analytics, but we included the initiative in the mapping because its aims strongly align with the other initiatives in this paper and because it was already mapped as a predictive analytics project by the Centre for Humanitarian Data.

[25] The initiative itself states that it is not predictive due to its inability to calculate risk of flooding of particular events. But we have included it anyway because its goals strongly align with other initiatives in the mapping and it had already been mapped by the Centre for Humanitarian Data.

activities and target disaster responses” (World Food Programme, 2020). Risk reduction activities linked to the initiative include Early Warning Early Action (EWEA), Shock Responsive Social Protection (SRSP), and Forecast-based Financing (FbF).

Where is flooding likely to occur? Google is developing a flood forecasting model that predicts where floods are likely to occur as well as predicting severity in order to generate early warning through Google public alerts (Nevo, 2019). Mercy Corps is piloting the use of drones in their ‘Managing Risk through Economic Development’ (M-RED) programme to monitor land-use changes and predict which areas are at risk of flooding in order to inform preventive measures (Sterling, 2017).

Where is disease outbreak most likely? Humanitarian Predictive Analytics Initiatives focusing on disease outbreaks focus primarily on predicting location. DFID, OCHA, UK Meteorological Office, NASA, University of Maryland and University of West Virginia have teamed up to predict which areas in Yemen are likely to experience a cholera outbreak in hopes to prevent outbreaks and support mitigation by those responding to outbreaks. The initiative has helped DFID and partners to focus resources in places where outbreaks are predicted; stock-pile prevention supplies and cholera treatment kits and medical equipment for hospitals (DFID, 2018). DFID claims this initiative may have contributed to a decrease of cholera cases in Yemen. After its implementation, “there were only 672 suspected cases of cholera in July 2018 compared to 13,659 in July 2017”. Similarly, the Global Cholera Risk Model, a proof of concept by the University of Florida, the University of Maryland and NASA, promises to predict outbreaks to a high degree of locational specificity (NASA, n.d.).

Where mosquito-transmitted diseases will occur is another area of focus for predictive analytics. UNICEF partnered with Google to predict Zika outbreaks in Latin America by region and hopes the model could be repurposed for future outbreaks of other mosquito-borne diseases (Bentley & Kerry, 2016). A start-up named Artificial Intelligence in Medical Epidemiology (AIME) suggests that it can predict mosquito-transmitted diseases at 400-metre resolution in real-time and has used this data to predict the location of disease outbreaks up to 3 months ahead of time with an 88.7% accuracy rate enabling prevention and mitigation measures (AIME, 2019). Premise Data and the local government in Cali, Colombia predict mosquito vector sites across the city to prevent outbreaks by identifying and destroying potential mosquito breeding grounds (Novak, 2018).

Where will there be conflict or atrocities? Initiatives seeking to answer this question mainly mentioned goals of prevention as their primary goal. The Simon-Skjodt Center and Dartmouth College’s ‘Early Warning Project’ seeks to predict which countries are at risk of mass killings in 160 countries in the next two years in hopes that early warning signs can be detected, addressed and mass killings prevented (Simon-Skjodt Center, n.d.). Similarly, ViEWS aims to predict “armed conflict involving states and rebel groups, armed conflict between non-state actors, and violence against civilians … [and] can analyse violence at national level, subnational level and actor level”. ViEWS hopes its early warning may help prevent, mitigate and adapt to large-scale political violence (Uppsala Universitet, n.d.).

Where will military rockets land? The Sentry project by start-up Hala Systems Inc. predicts where military planes in Syria are likely to attack 7 minutes in advance in time to provide early warning alerts to citizens via short message service (SMS), instant messaging, and sirens to prevent loss of life. According to Hala Systems’ preliminary analyses, the initiative has “resulted in an estimated 20-30% reduction in casualty rates in several areas under heavy bombardment in 2018” (Hala Systems Inc., 2019: 2). The Pakistan Safe Schools Initiative by start-up

Predictify.Me was an initiative that predicted the risk of schools being attacked and the losses they would suffer from an attack, including suicide bombings, and then provided preparedness options accordingly (Inge, 2015; Watt, 2015).[26]

Where will displaced people go? Brunel University's Flee project seeks to predict where forcefully displaced people will migrate so that humanitarian actors are better prepared for their arrival and more efficiently allocate their resources to accommodate them (Buchanunn, 2017). IBM and the Danish Refugee Council's Mixed Migration Foresight (MM4Sight) tool also seeks to predict migration flows, where people will migrate, and how they will get there in order to help prepare and plan more efficient responses (Danish Refugee Council & IBM, 2018; Novack, 2018).

Where will the situation be worst? Predicting where the most acute need will be was the focus of the Netherland Red Cross' 510 Typhoon Impact Model that seeks to predict which locations will be hardest hit by typhoons and hurricanes in order to mobilise resources optimally (Van Der Veen, 2016).

Predicting when

Although all predictive analytics is concerned with events in the future, not all predictive analytics is specifically focused on predicting the timing of those events. In our study, predictive analytics focused on determining 'when' events would occur accounted for just nine initiatives (18%). Understanding when the effects of a humanitarian emergency will be felt can be as important as knowing where it will be felt. Whether you are predicting the course of flooding from the highlands into the floodplains or the paths of famine or disease, timing is of the essence. The ability to (pre)deploy finance, supplies and personnel is a matter of timing as much as location.

When will it begin? The CALM initiative is able to make predictions about cholera outbreaks in specific municipalities across Yemen at multiple time intervals ranging from 2 weeks to 2 months. Including multiple intervals has proven especially useful the initiative because longer time-frames are unable to capture the emergence of sudden spikes that require immediate response, whereas longer time-frames provide humanitarian actors with the time needed to better prepare for future outbreaks through prevention and mitigation strategies (Badkundri et al., 2019). The Pakistan Safe Schools Initiative by start-up Predictify Me provided schools with a 3–7 day window warning of when they are likely to be subject to a terrorist attack. It was believed to be able to predict attacks 3 days in advance with 72% accuracy (Inge, 2015; Kavilanz, 2015). Similarly, Hala Systems provides citizens with an early warning of a few minutes about an imminent air strike (Hala Systems Inc., 2019).

When will it end? Save the Children's Migration and Displacement (MDI) initiative seeks to predict the duration and scale of displacement in order to better prepare responses and determine whether short-term humanitarian relief or long-term development solutions should be prioritised in response to displacement. The hope is that this will allow Save the Children to make an early case for funding and use funds more efficiently (Morgan & Kaplan, 2018).

8. Future Plans

This section looks at the future plans for actors in humanitarian predictive analytics. In the data that we had access to there was relatively little information provided about future plans. No future

[26] The company that ran this initiative now appears to be out of business.

plans were available for 22 of the initiatives (46%). Details available on many others were limited to short website descriptions or quotes in blogs or other media. The future plans that were identifiable fell into two main categories of increasing depth and increasing breadth (i) plans to enhance and improve their existing predictive models and (ii) plans to scale existing predictive models to new subjects, new geographies, and new organisations.

Improvements to existing predictive models were planned by several actors. Eight initiatives mentioned planning to feed additional data sources to their existing models. Two initiatives planned to explore the use of additional data collection technologies. Six organisations had plans to improve the predictive accuracy of their models. Two initiatives were planning to modify their models to enable them to make predictions further into the future. Two initiatives reported plans to automate some of the predictive processes. We also found one initiative wanting to improve the speed at which predictions could be made.

Improving the scale of existing predictive analytics featured in the plans of many agencies. Four initiatives reported plans to expand the scope of their work by predicting additional factors via their models. At least eight initiatives are planning to apply their predictions to new places or event types. Three organisations were already scaling efforts, two of which include efforts to scale initiatives through innovation uptake by other organisations. For example, OCHA has expressed interest in scaling DFID's project on cholera outbreak prediction (produced in partnership with the Met Office, NASA and others) to new contexts (CERF, 2019) and the Danish Refugee Council has now taken the lead role in scaling IBM's refugee prediction model. Eight initiatives stated that they were seeking more collaborations and partnerships.

Save the Children plans to build a cross-sector platform that allows data to be collected, shared, analysed, and incorporated into models that can better predict displacement patterns (Morgan & Kaplan, 2018). The WFP's PRISM initiative is looking into user research to understand how to better understand the needs of external organisations that may be potential users of their predictive model. Similarly, one proof-of-concept study by Lancaster Medical School, Liverpool School of Tropical Medicine, Malawi-Liverpool-Wellcome Trust Clinical Research Programme, and UNICEF looking into the use of drones to predict malaria outbreaks suggested they were seeking to make their solution as user-friendly as possible in order to improve the prospects of local ownership.

We did not find any future plans which mentioned the inclusion of affected populations in the design, implementation or evaluation of predictive analytics. This raises the question of accountability and the issue of 'Whose predictions?', 'Whose priorities?' and 'Whose interests are being served by predictive analytics in humanitarian aid?'. Future research could usefully investigate the range and relative successes of different pathways to scale for predictive analytics innovation in humanitarian action including franchising, open-sourcing, internal adoption and networked uptake.

9. Risks and ethics of predictive analytics

The terms predictive analytics, machine learning, data science and artificial intelligence can convey a false impression of scientific certitude. However, as in other areas, predictive analytics is based on biased data and incomplete models about dynamic and often chaotic situations. The resulting predictions are only probability calculations (with significant margins of error) about fundamentally unknowable futures (Heffernan, 2020). Humanitarians must be mindful that the use of digital data, algorithms and automated decision-making introduces new risks and responsibilities. The collection and storage of data on vulnerable populations and the production of predictions to guide practice raises new ethical issues of data safeguarding and informed consent that must be weighed against crucial humanitarian principles. This section details some of the key issues arising.

Not everything that is important can be predicted. Some crises are easier to predict than others. It is far easier to make reliable predictions in slow-onset disasters like droughts than it is to predict rapid onset disasters like earthquakes. There is an ongoing debate about whether rapid onset disasters like earthquakes can be predicted at all (Asencio–Cortés et al., 2018; Stierwalt, 2020). This may lead humanitarian actors to focus their predictive analytic efforts on the low-hanging fruit, the crises that are relatively easy to predict. Although predictive analytics may turn out to be a useful tool for mitigating, preparing and responding to crises, an over-reliance on predictive analytics may lead to the neglect of important but less predictable issues and an unintended silencing of those whose experiences are not represented in historical data or whose voices are not present on social media. *Humanitarian actors using predictive analytics must therefore be at pains to ensure that they remain people-centred and problem-driven rather than technology-centred and solution-driven.*

Change is constant. A major assumption made by those implementing predictive analytics is that the future will be like the past (Davenport, 2014). Although people tend to be 'creatures of habit' and maintain their behaviour patterns over time (Duhigg, 2013), predictive models may not be very good at picking up instances where behaviour is consciously changed, or when a big and rare event like the financial crises of 2008—or COVID-19—changes the opportunities and decisions available (Davenport, 2014). The more time that has passed since it was updated, the less accurate models are at prediction. This is illustrated by weather forecasts. Weather models can forecast the weather five days from now with 90% accuracy, but ten-day forecasts are only 50% accurate.[27]

***Why* things happen remains unclear.** Predictive models find patterns or correlations in data, but correlation is not causation. Understanding why an event happens is outside the scope of humanitarian predictive analytics. Correlations in big data allow statistical models to pattern and predict who, what, where and when events will happen – but not *why*. So even if predictions are accurate, we are left no wiser about their causes. Advocates of predictive analytics argue that knowing what will happen is more important than knowing why it happens if it enables us to take action to produce more desirable outcomes. "It just needs to work; prediction trumps explanation" argues Siegel (2016: 132). However, this limits human action to responding to the symptoms of a problem rather than identifying and tackling its root cause.

Bias in data. Data scientists have long used the maxim 'garbage in – garbage out' to describe the dependency of computer analysis outputs on the quality of the data inputs. Predictive analytics is entirely dependent on the quality of the data sets on which the machine learning and statistical models operate. This problem was exemplified when Amazon human resource management used machine learning to analyse all of its own historic employment appointments and then used them to predict and select candidates for interview. Their model reproduced their historical pattern of gender discrimination and disproportionately recommended appointing male candidates.[28] In this case, the maxim 'garbage in – garbage out' could be modified to 'patriarchy in – patriarchy out'. The example is relevant here not because we found humanitarians using

[27] https://scijinks.gov/forecast-reliability/

[28] https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

predictive analytics in their human resource management (although we did) but because machine learning trained on patterns and correlations in the global North may perform poorly in other contexts. For example, a winning entry to a machine learning competition aiming to detect buildings from satellite imagery achieved 89% accuracy in Las Vegas, but only 42% in Khartoum, Sudan (Paul et al., 2018). In other cases, facial recognition software has been shown to be less likely to recognise darker faces (Benjamin, 2019).

Data gaps. The experiences and realities of the demographic groups that humanitarian action seeks to serve are not well represented in commonly available data sets. “Data related to development challenges are often scarce and difficult to obtain, especially in a digital format conducive to machine learning” (Paul et al., 2018: 44). Existing data sets over-represent white men from the global North and make invisible black women from the global South (Lerman, 2013; Noble, 2018; Criado Perez, 2019). The structurally low levels of representation of marginalised people in data sets are, in part, a function of their relative exclusion from digital connectivity, ownership and use. For example, predictive analytics increasingly draws on social media data to gain situational awareness in humanitarian settings. However, social media users are a relatively privileged demographic and attention to their data may make invisible the needs and priorities of more vulnerable affected populations. Over-reliance on social media data risks humanitarian organisations falling victim to a ‘signal problem’ where people in areas within a country that are less connected risk not having their realities reflected in the data. Moreover, the same may be true between richer and poorer parts of highly unequal cities (Crawford & Finn, 2015). Using crisis maps based on this data may lead to assessments showing that the more affluent areas are the worst impacted by a disaster while making the situations of the poorest and most vulnerable invisible (Sharma & Joshi, 2019).

Reproducing (dis)advantage. We live in societies which contain structural inequalities including ones of gender, race, class and caste. Machine learning that is trained on data from unequal societies tends to reflect, reproduce and amplify those inequalities (Benjamin, 2019). Predictive analytic models do not understand the meaning of the data that they sift through – they just find patterns and use them to predict futures. Predictive analytic models also do not have any sense of concepts such as justice or fairness or an understanding of historical injustice. Predictive analytics used in bail decisions that use historical data has been shown to reproduce racial (dis)advantage (Benjamin, 2019). Predictive analytics in recruitment settings that uses historical hiring data has been shown to reproduce gender (dis)advantage (Paul et al., 2018). The risk, therefore, exists that predictive analytics in humanitarian action will have the unintended consequence of reproducing historical patterns of discrimination and inequities. A workshop held by the Centre for Humanitarian Data drew attention to these risks.

> The tendency of algorithmic methods to learn and codify the current state of the world and thereby make it harder to change, seemed of particular concern to participants in relation to current use cases for predictive analytics in the humanitarian sector, e.g. are we building models that perpetuate and encode past mistakes? (Centre for Humanitarian Data, 2019a: 7).

These concerns were also echoed in the recent workshop on the use of artificial intelligence in humanitarian aid convened by UK Research and Innovation. Participants reflected on the danger of automating inequalities and dehumanising humanitarian processes. They concluded that there is an important role for researchers to work with affected populations and humanitarian practitioners to build frameworks that align with the precautionary principle to ‘do no harm’ (Roberts & Faith, in press).

Data resilience, viability and veracity after disasters. Physical digital infrastructures are often destroyed in disasters, and targeted in conflict, such that reliance on digital data can create blind spots. Infrastructure destruction may make areas virtually invisible (Sharma & Joshi, 2019). Social media may allow humanitarian actors to collect data very quickly after a disaster but making sense of this data can be quite difficult. The data is often messy, in multiple local languages and the shared content often includes highly contextual meanings which can be difficult to decode, especially by foreign data scientists. Moreover, digital data collected after the onset of a disaster can contain a lot of noise, especially when it is crowdsourced (Sharma & Joshi, 2019). There is also the possibility of inauthentic digital actors (trolls, bots or cyborgs) intentionally spreading false information (Bradshaw & Howard, 2018).

Data privacy. To counter the risks of data misuse it is often proposed that data is anonymised. However, it has been proven that by linking and combining multiple data sets it is possible to de-anonymise data in what is known as the mosaic effect (Mazmanian, 2014). Moreover, ethical issues of informed consent arise when individuals provide permission for their data to be used for one purpose (e.g. a pandemic) but it is then shared for other purposes without the individual's informed consent for that purpose (UN Global Pulse, 2017). In many cases, the data used to make predictions is secondary historical data—collected for a purpose other than prediction—and in some cases pre-exists the widespread use of machine learning, an application which might have been unimaginable for some of those consenting to have their data collected (Ballantyne, 2018).

Accountability. The humanitarian sector has long struggled with issues of accountability to affected populations (Alexander et al., 2013). Predictive analytics adds to this problem. Machine learning algorithms are a 'black box' meaning that their internal workings are most often inaccessible or unknowable. Machine learning finds patterns without precise instruction and not even the people who make truly understand *why* they make particular predictions. This presents ethical dilemmas about transparency and accountability. It can be argued that machine learning is dehumanising by design as it aims to replace human dialogue and deliberation with automated machine calculation. As such algorithmic decision-making may be in potential conflict with humanitarian commitments to human-centred and participatory processes (Roberts & Faith, in press). Sharma and Joshi (2019) argue that 'digital humanitarians' working online in other countries tend not to have local contextual knowledge of disaster or humanitarian settings and may, therefore, be less able to assess whether their work will 'do no harm' (Sharma & Joshi, 2019).

The sector has recognised a need to put in place guidelines and governance mechanisms to make sure that predictive analytics does not cause harm. The Centre for Humanitarian Data is active in this area organising a peer review mechanism. There is also a new Data Science and Ethics Group[29] which has published a Framework for the Ethical Use of Advanced Data Science Methods in Humanitarian Sector (DSEG, 2020). The Centre for Humanitarian Data have reminded practitioners of the value of keeping 'humans in the loop' and cautioned that predictive analytic models should be seen as tools rather than solutions – which require the involvement of human decision-makers from the beginning in order to be successful (Centre for Humanitarian Data, 2019a).

[29] https://datascienceinitiative.eu/events/introducing-the-humanitarian-data-science-and-ethics-group-dseg-why-do-we-need-an-ethical-framework/

10. Conclusion

This report has provided a preliminary mapping and analysis of humanitarian predictive analytics projects. Despite the limitations of a rapid response review, this study is the most extensive mapping and analysis to date of predictive analytics applied to humanitarian action. Although further research is necessary to extend the breadth of initiative covered and to verify the detail of these findings, we have been able to map the principle contours and features in the emerging landscape of humanitarian predictive analytics.

This study has provided evidence that predictive analytics is being employed to a wide range of humanitarian problems in a rapidly growing number of applications. The potential for expansion to new humanitarian problems, geographies and sectors is great. Currently, predictive analytics is mainly focused on humanitarian problems in the areas of disease outbreak, conflict, migration, disaster risk reduction, and food security, with lower levels of activity in other operational areas as well as key support functions of fundraising, human resources and logistics. The study has found that humanitarian predictive analytics is mainly being used to predict where humanitarian events will happen (71%), who will be most affected (40%), what the key features of the humanitarian emergency will be (26%) and when events are most likely to occur (18%). Geographically, the initiatives mapped are concentrated in Africa and the Arab world with lower levels of activity in Asia and Latin America. Technically, predictive analytics uses a wide range of novel data sources including open data, social media data, and satellite and drone images in order to produce the big data necessary for machine learning and statistical modelling. The most common technical approaches being employed include natural language processing, image recognition and statistical modelling using simulations and scenarios.

The ecosystem of predictive analytics is developing slowly with the Centre for Humanitarian Data playing a key convening role and other sectoral and geographical specialist agencies emerging. The use of predictive analytics in humanitarian work remains in its infancy and currently, it is being used alongside - not in place of - traditional humanitarian forecasting and early-warning systems. Our study shows that predictive analytics is carried out predominantly by large international humanitarian organisations and commercial start-ups and that partnership between multiple actors are common. However, the specialised skill sets and high cost of data science expertise make it very difficult for indigenous and small local agencies to lead on applications of predictive analytics. We found little evidence that affected populations are being engaged in predictive analytics projects by humanitarian agencies.

Limited time and reliance on secondary data meant that we had incomplete data which made it difficult to determine whether initiatives were ongoing or to verify claims made. Reliance on secondary data meant that we were able to gain limited insight into the risks, ethical dilemmas, future plans or barriers to scale experienced by predictive analytics actors. Primary research is necessary to illuminate these gaps in knowledge in this important and rapidly developing area. Although it is clear that partnerships are common in predictive analytics initiatives, no information was available about the motivation, relative contributions, tension or benefits of the partnerships. We were not able to find sufficient information on sources of finance for humanitarian predictive analytics to present any findings. Primary research would be necessary to ascertain these details.

The collection and sharing of data on vulnerable populations and its use to inform humanitarian action presents serious ethical issues and risks. The desire to innovate, advance humanitarian response and gain efficiencies need to be balanced against humanitarian principles. Predictive analytics is not an exact science. It uses biased data and incomplete models to calculate probabilities. Humanitarians must be mindful that the use of digital data, algorithms and automated decision-making are only tools to be used alongside the grounded knowledge of affected populations and experience of practitioners.

Recommendations

Our mapping and analysis give rise to four preliminary recommendations: the need to engage affected populations, strengthen the ecosystem, attend to risks, and conduct primary research in order to deepen understanding of the current situation and future directions.

Affected populations: There is an opportunity at this early stage to actively engage affected populations in the design, implementation and evaluation of humanitarian predictive analytics. The lessons from participatory development and user-centred design can usefully inform an iterative action research project cycle in which the contextual knowledge of affected populations and the experience of humanitarian practitioners are combined with the technical expertise of data scientists to improve both the power relationships and predictive efficacy of future innovations in predictive analytics.

Ecosystem development: The ecosystem of humanitarian predictive analytics is emerging with effective leadership from the Centre for Humanitarian Data, OCHA and the Humanitarian Data Science and Ethics Group (DSEG). Funders should support the development of ecosystem diversity by strengthening organisations with sectoral specialisms like Precision Public Health or with geographic specialisms located in the global South like members of Big Data for Development. The promotion of open data, the sharing of lessons, tools and standards, and partnerships that enable dissemination, adoption and uptake are key.

Risks and downsides: Humanitarian agencies must apply the precautionary principle in data collection, data safeguarding, and responsible data to protect vulnerable populations from harm. The pressure to innovate, advance humanitarian action and secure efficiencies should never lead to experimenting on vulnerable populations. Current data security and data protection practices need significant improvement. A dialogue needs to take place within the humanitarian sector about the tension between human-centred processes and algorithmic decision-making.

Further research: Our knowledge about this crucial and rapidly changing area would be greatly improved by primary research. Interviews with leading innovators and experts would deepen our understanding of what is current, what is working, and what is planned. We know little about pathways to scale predictive analytics. A series of short well-documented case studies would help attract resources, facilitate dialogue, and stimulate innovation. Focus groups or practitioner workshops would help surface pinch points and dilemmas not publicly discussed.

References

African Risk Capacity (2019). *Africa RiskView: A software package for the assessment of impacts of natural hazards and supporting effective disaster response*. Johannesburg, South Africa: African Risk Capacity. https://www.africanriskcapacity.org/wp-content/uploads/2019/05/Africa-RiskView-brochure-May-2019.pdf

AIME (2019, November 13-15). *The use of AI in Dynamic Dengue Outbreak Surveillance & Forecasting*. New Delhi, India. https://www.itu.int/en/ITU-T/Workshops-and-Seminars/ai4h/201911/Documents/S2_Helmi_Zakariak_Presentation.pdf

Akter, S. & Wamba, S.F. (2019). Big data and disaster management: a systematic review and agenda for future research. *Annals of Operations Research* 283(1), 939–959. http://doi.org/10.1007/s10479-017-2584-2

Alexander, J., Darcy, J. & Kiani, M. (2013). *2013 Humanitarian Accountability Report.* Humanitarian Accountability Partnership International.

Army Technology (2015, November 8). *Safer with data: protecting Pakistan's schools with predictive analytics*. Army Technology. https://www.army-technology.com/features/featuresafer-with-data-protecting-pakistans-schools-with-predictive-analytics-4713601/

Asencio–Cortés, G., Morales–Esteban, A., Shang, X. & Martínez–Álvarez, F. (2018). Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure. *Computers & Geosciences* 115: 198–210. http://doi.org/10.1016/j.cageo.2017.10.011

Badkundri, R., Valbuena, V., Srikusmanjali, P., Cantrell, B. & Standeven, J. (2019, February 16). *Forecasting the 2017-2018 Yemen Cholera Outbreak with Machine Learning*. ArXiv:1902.06739. https://arxiv.org/abs/1902.06739

Ballantyne, A. (2018). Where is the human in the data? A guide to ethical data use. *GigaScience* 7.7. http://doi.org/10.1093/gigascience/giy076

Bastani, H., Bastani, O. & Kim, C. (2018). Interpreting Predictive Models for Human-in-the-Loop Analytics. *ArXiv Preprint ArXiv:1705.08504*: 45

Benjamin, R. (2019). *Race after technology: abolitionist tools for the new Jim code*. Polity

Bentley, A. & Kerry, F. (2016, March 3). *Google says its engineers working with UNICEF to map* Zika. Reuters. https://www.reuters.com/article/us-health-zika-alphabet/google-says-its-engineers-working-with-unicef-to-map-zika-idUSKCN0W50OR

BKV (2013). *BKV and American Red Cross to Present: Using Analytics to Mine Donor Data*. BKV Blog. http://bkv.bkvdigital.com/blog/bkv-and-american-red-cross-to-present-using-analytics-to-mine-donor-data/

Bradshaw, S. & Howard, P. (2018). *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation*. Working Paper 2018.1. Oxford, UK: Project on Computational Propaganda. comprop.oii.ox.ac.uk

Buchanunn, J. (2017, October 17). *New simulation technology to predict refugee destinations could improve aid efforts*. Brunel University London. Brunel University London. https://www.brunel.ac.uk/news-and-events/news/articles/New-simulation-technology-to-predict-refugee-destinations

Cannon, T. (2008). Reducing people's vulnerability to natural hazards communities and resilience. WIDER Research paper, No. 2009/34. http://hdl.handle.net/10419/45089

Castellucia, C. & Le Métayer, D. (2019, June 12). *Understanding algorithmic decision-making: opportunities and challenges.* Publications Office of the EU. http://publications.europa.eu/publication/manifestation_identifier/PUB_QA0618337ENN

Centre for Humanitarian Data (2019a, October 2). *Predictive analytics in humanitarian response*. Data Science Initiative. https://datascienceinitiative.eu/projects/predictive-analytics-for-humanitarian-response/

Centre for Humanitarian Data (2019b, April 18). *We are all Data People: Insights from the Data Literacy Survey*. Centre for Humanitarian Data. https://centre.humdata.org/we-are-all-data-people-insights-from-the-data-literacy-survey/

Centre for Humanitarian Data (2020). *Catalogue of predictive models in the humanitarian sector*. Centre for Humanitarian Data. https://centre.humdata.org/catalogue-for-predictive-models-in-the-humanitarian-sector/

CERF (2019). *CERF and Anticipatory Action: Internal Note*. United Nations Central Emergency Response Fund. https://cerf.un.org/sites/default/files/resources/CERF_and_Anticipatory_Action.pdf

CEWARN (n.d.). *Conflict Early Warning and Response Mechanism Brochure*. CEWARN. http://www.cewarn.org/attachments/article/51/CEWARN_Brochure.pdf

Chandir, S.; Siddiqi, D.A.; Hussain, O.A.; Niazi, T.; Shah, M.T.; Dharma, V.K.; Habib, A. & Khan, A.J. (2018). Using Predictive Analytics to Identify Children at High Risk of Defaulting From a Routine Immunization Program: Feasibility Study. *JMIR Public Health and Surveillance 4*(3):e63, http://doi.org/10.2196/publichealth.9681

Crawford, K. & Finn, M. (2015). The limits of crisis data: analytical and ethical challenges of using social and mobile data to understand disasters. *GeoJournal 80*(4): 491–502. http://doi.org/10.1007/s10708-014-9597-z

Criado Perez, C. (2019). *Invisible women: exposing data bias in a world designed for men*. Chatto & Windus.

Crisis Computing team at Qatar Computing Research Institute (2017, February 22). *About*. Crisis Computing. https://crisiscomputing.qcri.org/about/

Danish Refugee Council & IBM (2018). *Mixed Migration Foresight: Danish Refugee Council and IBM project on predictive modelling*. Mixed Migration Centre. http://www.mixedmigration.org/wp-content/uploads/2018/07/MM4Sight_1pager.pdf

Davenport, T. (2014). *A Predictive Analytics Primer*. Harvard Business Review. https://hbr.org/2014/09/a-predictive-analytics-primer

Davenport, T.H. (2015). 5 Essential Principles for Understanding Analytics. *Harvard Business Review*, 21 October. https://hbr.org/2015/10/5-essential-principles-for-understanding-analytics

DFID (2018, August 28). *World first as UK aid brings together experts to predict where cholera will strike next*. Department for International Development. https://hbr.org/2015/10/5-essential-principles-for-understanding-analyticshttps://hbr.org/2015/10/5-essential-principles-for-understanding-analytics

DSEG (2020, April 1). *A Framework for the Ethical Use of Advanced Data Science Methods in Humanitarian Sector.* Humanitarian Data Science and Ethics Group. https://5f2cd2ba-741c-4b29-ae47-00a8291b1d3c.filesusr.com/ugd/d1cf5c_6af8feb771194453817d62c92cee2a21.pdf

Duhigg, C. (2013). *The power of habit: why we do what we do and how to change*. Random House Books

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press

FEWSNET (n.d.). *Scenario Development*. Famine Early Warning Systems Network. https://fews.net/our-work/our-work/scenario-development

Fox, A. (2019, April 24) *Analytics at the American Red Cross*. Opex Analytics Blog, Medium. https://medium.com/opex-analytics/analytics-at-the-american-red-cross-6c12443f6e1

Gates Foundation (n.d.). *Integrated Platform to Identify Malaria Data "Cold-Spots"*. Grand Challenges. https://training.fema.gov/emiweb/downloads/is111_unit%204.pdfhttps://training.fema.gov/emiweb/downloads/is111_unit 4.pdf

Ginnetti, J. & Franck, T. (2014). *Assessing drought displacement risk for Kenyan, Ethiopian and Somali pastoralists*. Internal Displacement Monitoring Centre & Norweigan Refugee Council. http://doi.org/10.1163/2210-7975_HRD-9806-2014039

Haigh, R. (n.d.) *Disaster Management Lifecycle*. Learning Package 3: Disaster management lifecycle. University of Salford. http://www.science.earthjay.com/instruction/HSU/2017_spring/GEOL_308/lectures/lecture_01/GEOL_308_suppl_reading_02_Introduction_to_Disaster_Management_Lifecycle.pdf

Hala Systems Inc. (2019). *Hala Systems Inc.: Protect Everything that Matters*. https://drive.google.com/file/d/1KDGJnt8_wC_ymIK_GfjG3LfTogD22nQA/view

Heffernan, M. (2020). *Uncharted: how to map the future together*. Simon & Schuster

Hegre, H., Allansson, M., Basedau, M., Colaresi, M., Croicu, M. & Vestby, J. (2019). ViEWS: A political violence early-warning system. *Journal of Peace Research 56*(2), 155–174. http://doi.org/10.1177/0022343319823860

Hernandez, K. & Roberts, T. (2018). *Leaving No One Behind in a Digital World: An Extended Literature Review*. K4D Emerging Issues Report 10. Brighton, UK: Institute of Development Studies. https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/14147

Humanity Data Systems (2019). *HOPE (Humanitarian Operations Planning Environment)'.* Humanity Data Systems. http://humanitydatasystems.com/hope/

IDMC (n.d.). *Monitoring platform: How IDMC monitors*. Internal Displacement Monitoring Centre. https://www.internal-displacement.org/monitoring-tools/monitoring-platform

IFRC (2020). *Red Cross releases funds in anticipation of extreme winter in Mongolia*. International Federation of Red Cross and Red Crescent Societies. https://media.ifrc.org/ifrc/press-release/red-cross-releases-funds-anticipation-extreme-winter-mongolia/

Inge, L. (2015, March 30). *Raleigh's Predictify Me; To Predict Suicide Bombings in Pakistan*. North Carolina Public Radio. https://www.wunc.org/post/raleighs-predictify-me-predict-suicide-bombings-pakistan

International Crisis Group (2016, July 28). *About CrisisWatch.* International Crisis Group. https://www.crisisgroup.org/about-crisiswatch

International Red Cross and Red Crescent Movement (2019). *Forecast-based Financing in the Red Cross Red Crescent Movement*. https://drive.google.com/file/d/1T6Z1bO1uvtRvSFaCkVtSXCw1IQXFDVQO/view

JADE (n.d.). *Regional Partnership to Support Efficient Humanitarian Response.* Pacific Disaster Center. https://www.pdc.org/wp-content/uploads/PDC-WFP-UNOCHA-Partnership.pdf

Kaplan, E. (2017, October 26). *"Smart Refugee Camps": applying the best of IoT and ICT for better camp management*. UNICEF Jordan Blog, Medium. https://medium.com/@unicefjordan1/smart-refugee-camps-applying-the-best-of-iot-and-ict-for-better-camp-management-e35e619e7310

Kavilanz, P. (2015, May 1). *Startup PredictifyMe could help thwart school bombings*. CNN Business. https://money.cnn.com/2015/03/26/smallbusiness/predictifyme-school-bombings/index.html

Lerman, J. (2013). Big Data and Its Exclusions. *Stanford Law Review Online 66*(55). https://review.law.stanford.edu/wp-content/uploads/sites/3/2016/08/66_stanlrevonline_55_lerman.pdf

Letouze, E., Sangokoya, D. & Ricard, J. (2017). Big Data: Predicting and preventing climate related shocks. *Integration & Trade Journal* 21.41

Lu, X., Bengtsson, L. & Holme, P. (2012). Predictability of population displacement after the 2010 Haiti earthquake. *Proceedings of the National Academy of Sciences 109*(29) 11576–11581. http://doi.org/10.1073/pnas.1203882109

Mazmanian, A. (2014, May 13). *The mosaic effect and big data.* The Business of Federal Technology. https://fcw.com/articles/2014/05/13/fose-mosaic.aspx

McKechnie, C. & Axelson, D. (Producers). (2019, October 10). *AI in healthcare and pandemics* [Audio podcast]. Discovery Matters. https://www.gelifesciences.com/en/us/news-center/ai-in-healthcare-and-pandemics-10001

Milano, L. (2017, January). *#IDETECT: how technology and collaboration between innovators can help ensure "no one is left behind"*. Internal Displacement Monitoring Centre (IDMC)

Blog. https://www.internal-displacement.org/expert-opinion/idetect-how-technology-and-collaboration-between-innovators-can-help-ensure-no-one

Morgan, S. & Kaplan, J. (2018). *Predicting Displacement: Using Predictive Analytics to Build a Better Future for Displaced Children*, Save the Children International. https://resourcecentre.savethechildren.net/library/predicting-displacement-using-predictive-analytics-build-better-future-displaced-children

MSF (2018). *People Analytics Summary*. Médecins Sans Frontières Transformational Investment Capacity (TIC). http://msf-transformation.org/wp-content/uploads/2018/11/People-Analytics-Summary-2018.09.27.pdf

NASA (n.d.). *Using Precipitation Data to Map Cholera Risk*. Precipitation Measurement Missions. https://gpm.nasa.gov/articles/using-precipitation-data-map-cholera-risk

Nevo, S. (2019, September 18). *An Inside Look at Flood Forecasting.* Google AI Blog. http://ai.googleblog.com/2019/09/an-inside-look-at-flood-forecasting.html

Noble, S.U. (2018). *Algorithms of oppression: how search engines reinforce racism*. New York University Press

Novack, R. (2018, June 20). *Technology could help predict the next refugee crisis — here's how*. Business Insider. https://www.businessinsider.com/sc/how-to-predict-the-next-refugee-crisis-2018-6

Novak, M. (2018). *Fighting Zika with Digital Vector Surveillance in Cali, Colombia.* Premise. https://www.premise.com/fighting-zika-with-digital-vector-surveillance-cali-colombia/

OCHA (2018, September 24). Famine Action Mechanism: Predictive data, faster financing and better joint response to rid the world of the scourge of famine. United Nations Office for the Coordination of Humanitarian Affairs https://www.unocha.org/story/famine-action-mechanism-predictive-data-faster-financing-and-better-joint-response-rid-world

OCHA (2019a). *Global Humanitarian Overview 2019*. Humanitarian Response. https://www.humanitarianresponse.info/sites/www.humanitarianresponse.info/files/documents/files/gho-2019-digital-04122019_iii.pdf

OCHA (2019b, September 24). The Centre's Predictive Analytics Focus: Models, Peer Review, Community - World. Relief Web. https://reliefweb.int/report/world/centres-predictive-analytics-focus-models-peer-review-community

O'Donnell, A. (2015). *A Rights Based Approach to Responsible Data*. Policy and Practice Blog, Oxfam. http://policy-practice.oxfam.org.uk/blog/2015/08/a-rights-based-approachto-treating-data-responsibly

O'Neil, C. (2017). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Penguin Books

Paul, A., Jolley, C. & Anthony, A. (2018). *Reflecting the Past, Shaping the Future: Making AI Work for International Development*. USAID. https://www.usaid.gov/sites/default/files/documents/15396/AI-ML-in-Development.pdf

REACH (2019). *Yemen Flood Susceptibility Analysis Methodology*. Accessed April 2020.

Roberts, T. & Faith, B. (In press). *Digital Aid: understanding the digital challenges facing humanitarian assistance*, A GCRF Research Practitioner Workshop Report, Swindon, UK: United Kingdom Research and Innovation.

Roberts, T. & Marchais, G. (2017). *Assessing the Role of Social Media and Digital Technologies in Violence Reporting*. IDS Working Paper 492. Brighton: Institute of Development Studies. https://www.ids.ac.uk/publications/assessing-the-role-of-social-media-and-digital-technology-in-violence-reporting/

Rockefeller Foundation (2019). *The Rockefeller Foundation's Precision Public Health Initiative*. https://www.rockefellerfoundation.org/wp-content/uploads/Precision-Public-Health-Overview.pdf

Sharma, P. & Joshi, A. (2019). Challenges of using big data for humanitarian relief: lessons from the literature. *Journal of Humanitarian Logistics and Supply Chain Management* ahead-of-print (ahead-of-print). http://doi.org/10.1108/JHLSCM-05-2018-0031

Siegel, E. (2016). *Predictive analytics: the power to predict who will click, buy, lie, or die*. Wiley

Simon-Skjodt Center (n.d.). *Methodology: Statistical Model*. Early Warning Project. https://earlywarningproject.ushmm.org/methodology-statistical-model

Stanton, M. & Jones, C. (2018, July 6). *How Drones Are Helping in the Fight Against Malaria*. Global Citizen. https://www.globalcitizen.org/en/content/drones-malaria-unicef-malawi/

Sterling, T. (2017, December 12). *The NGO's Eye in the Sky.* Taun Sterling Blog, Medium. https://medium.com/@tsterl20/the-ngos-eye-in-the-sky-d6ec520dd4b0

Stierwalt, S. (2020, January 15). Can We Predict Earthquakes At All?. Scientific American. https://www.scientificamerican.com/article/can-we-predict-earthquakes-at-all1/

Swaminathan, J.M. (2018). Big Data Analytics for Rapid, Impactful, Sustained, and Efficient (RISE) Humanitarian Operations. *Production and Operations Management 27*(9). 1696–1700. http://doi.org/10.1111/poms.12840

Team Lambert (2018). *CALM Model*. CAPTIVATE. http://2018.igem.org/Team:Lambert_GA/CALM_MODEL

UN Global Pulse (2017). *Big Data for Development and Humanitarian Action: Towards Responsible Governance.* UN Global Pulse. https://www.unglobalpulse.org/wp-content/uploads/old_site/Big_Data_for_Development_and_Humanitarian_Action_Report_Final_0.pdf

UNHCR (2019, May 14). Is it possible to predict forced displacement? UNHCR Innovation Service Blog, Medium. https://medium.com/unhcr-innovation-service/is-it-possible-to-predict-forced-displacement-58960afe0ba1

Uppsala Universitet (n.d.). *About ViEWS.* Uppsala Universitet. https://www.pcr.uu.se/research/views/about-views/

Van Der Veen, M. (2016, November 9). *Efficient Aid: Through Accurate Priority Index Part 1*. 510 GLOBAL. https://www.510.global/philippines-typhoon-haima-priority-index/

Watson, H., Finn, R.L. & Wadhwa, K. (2017). Organizational and Societal Impacts of Big Data in Crisis Management. *Journal of Contingencies and Crisis Management 25*(1). 15–22. http://doi.org/10.1111/1468-5973.12141

Watt, E. (2015, March 18). *Technology initiative to help make 1000 schools safer in Pakistan*. Theirworld. https://theirworld.org/news/technology-initiative-to-help-make-1000-schools-safer-in-pakistan

World Food Programme (2020, January 31). *PRISM: Real-time impact and situation monitoring.* World Food Programme. https://innovation.wfp.org/project/prism

Appendix: List of initiatives included

Initiative	Organisation(s)	Field	Countries	Link
510 Typhoon impact model	Netherlands Red Cross - 510	Disaster Risk Reduction	Philippines	https://www.510.global/philippines-typhoon-haima-priority-index/
African Risk Capacity	34 African Union member states, African Risk Capacity Agency, and ARC Insurance Company Limited	Disaster Risk Reduction	34 African Union countries	https://www.africanriskcapacity.org/
Artemis Model[30]	World Bank, UN, ICRC, Microsoft, Google, Amazon Web Services	Food security	Currently in five countries, to be expanded to 21 more "Somalia, along with South Sudan, Afghanistan, Niger, and Mali are likely pilots for the data analysis."	https://www.worldbank.org/en/programs/famine-early-action-mechanism
Artificial Intelligence in Medical Epidemiology (AIME)	AIME	Disease outbreaks	Rio de Janeiro, Singapore, and two different states in Malaysia	https://www.itu.int/en/ITU-T/Workshops-and-Seminars/ai4h/201911/Documents/S2_Helmi_Zakariak_Presentation.pdf

[30] Also known as Famine Action Mechanism.

Cholera Artificial Learning Model (CALM)	Team Lambert	Disease outbreaks	Yemen	http://2018.igem.org/Team:Lambert_GA/CALM_MODEL
Cholera Prevention in Yemen	UKAID, OCHA, UK Meteorological Office, NASA, University of Maryland and University of West Virginia	Disease outbreaks	Yemen	https://www.gov.uk/government/news/world-first-as-uk-aid-brings-together-experts-to-predict-where-cholera-will-strike-next
Conflict Early Warning and Response Mechanism (CEWARN)	Intergovernmental Authority on Development (IGAD), mention of diverse partners	Conflict	Djibouti, Eritrea, Ethiopia, Kenya, Somalia, South Sudan, Sudan, Uganda	http://www.cewarn.org/attachments/article/51/CEWARN_Brochure.pdf
Crisis Computing/ Artificial Intelligence for Digital Response	Qatar Computing Research Institute, UN OCHA, UNICEF, Suffolk County Fire Rescue and Emergency Management Department in New York, and Education Above All.	Conflict	Global	https://crisiscomputing.qcri.org/about/
CrisisWatch	The International Crisis Group	Conflict	80 countries	https://www.crisisgroup.org/crisiswatch

Drones vs. Mosquitos[31]	Lancaster Medical School (Lancaster University), Liverpool School of Tropical Medicine, Malawi-Liverpool-Wellcome Trust Clinical Research Programme, UNICEF	Disease outbreaks	Malawi	https://www.designindaba.com/articles/creative-work/how-drones-are-helping-fight-against-malaria
Famine Early Warning Systems Network (FEWSNET)	USAID, NASA, NOAA, USDA, USGS	Food security	28 countries with focus on west, southern and east Africa	https://fews.net/nosso-trabalho
Fighting Zika with Digital Vector surveillance	Premise Data, Local Government	Disease outbreaks	Cali (Colombia)	https://www.premise.com/fighting-zika-with-digital-vector-surveillance-cali-colombia/
Flee	Brunel University London	Displacement and migration	Three African countries	https://www.brunel.ac.uk/news-and-events/news/articles/New-simulation-technology-to-predict-refugee-destinations
Forecast-based Financing in the International Red Cross and Red Crescent Movement	International Red Cross and Red Crescent Movement, information flows to many	Disaster Risk Reduction	Global	https://drive.google.com/file/d/1T6Z1bO1uvtRvSFaCkVtSXCw1IQXFDVQO/view

[31] A project in UNICEF's humanitarian drone testing corridor.

	different red cross organisations and others[32]			
Global Cholera Risk Model (GCRM)	University of Florida, University of Maryland, NASA	Disease outbreaks	Global	https://gpm.nasa.gov/articles/using-precipitation-data-map-cholera-risk
Global Disaster Displacement Risk Model	IDMC, UNDRR, ETH Zurich	Displacement and migration	Global	https://www.internal-displacement.org/disaster-risk-model
Humanitarian Enterprise Logistics and Provisioning (HELP)	Humanity Data Systems	Logistics	Focus on the Middle East	http://humanitydatasystems.com/help-2/
Humanitarian Operations Planning Environment (HOPE)	Humanity Data Systems	Disaster Response	Focus on the Middle East	http://humanitydatasystems.com/hope/
Hunger Map Live	WFP and Alibaba	Food security	Global	https://hungermap.wfp.org/
IBM's Refugee & Migration Predictive Analytics Solution prototype	IBM	Displacement and migration	Syria	https://www.businessinsider.com/sc/how-to-predict-the-next-refugee-crisis-2018-6?r=US&IR=T

[32] WMO, ODI, Frankfurt School UNEP collaboration centre, and an additional organisation whose name is not legible in the publication.

Improving the collection of food aid[33]	Catholic Relief Services, Esri Professional Services	Food security	Madagascar	https://www.esri.com/about/newsroom/arcuser/predictive-analysis-brings-food-aid-closer-in-madagascar/
Integrated Platform to Identify Malaria Data "Cold-Spots"	Dimagi, The Arnhold Institute for Global Health	Disease outbreaks	Senegal, Gambia	https://gcgh.grandchallenges.org/grant/integrated-platform-identify-malaria-data-cold-spots
Methods for Extremely Rapid Observation of Nutritional Status (MERON)	Kimetrica	Nutrition	Kenya	https://kimetrica.com/our-projects/
Migration and Displacement Initiative (MDI)	Save the Children	Displacement and migration	Global	https://resourcecentre.savethechildren.net/node/14290/pdf/predicting_displacement_report_-_save_the_children_mdi.pdf
Mixed Migration Foresight Project	Danish Refugee Council and IBM	Displacement and migration	Afghanistan, Myanmar	http://www.mixedmigration.org/wp-content/uploads/2018/07/MM4Sight_1pager.pdf
Modelling Early Risk Indicators to Anticipate Malnutrition (MERIAM)	Action Against Hunger, The graduate institute of Geneva, Johns Hopkins University, University of Maryland	Nutrition	Kenya, Niger, Nigeria, Somalia and Uganda	https://www.actionagainsthunger.org/meriam

[33] Not the initiative's official name, no name provided by the organisation.

Mongolia dzud model	National Agency for Meteorology and Environmental Monitoring (NAMEM), IFRC	Livelihoods	Mongolia	https://media.ifrc.org/ifrc/press-release/red-cross-releases-funds-anticipation-extreme-winter-mongolia/
Optimizing local volunteer deployment	American Red Cross	Human resources	Chicago, USA	https://medium.com/opex-analytics/analytics-at-the-american-red-cross-6c12443f6e1
Pakistan Safe Schools Initiative[34]	PredictifyMe, UNICEF	Conflict	Pakistan	No link available[35]
Pastoralist Livelihood and Displacement Simulator	IDMC, Climate Interactive	Displacement and migration	Border regions of Kenya, Ethiopia and Somalia	https://www.internal-displacement.org/sites/default/files/publications/documents/201405-horn-of-africa-technical-report-en.pdf
People Analytics	MSF	Human resources	Global	http://msf-transformation.org/wp-content/uploads/2018/11/People-Analytics-Summary-2018.09.27.pdf

[34] Safe Schools Initiative is a larger UNICEF fund.

[35] The company's website appears to no longer exist. CrunchBase suggests the company closed its doors in 2018.

Predicting Demographic Trends for Global UNHCR Persons of Concern	Johnathon Shapiro[36]	Displacement and migration	Global	https://towardsdatascience.com/predicting-demographic-trends-for-global-unhcr-persons-of-concern-86dc4b8b920d
Predicting who will donate	American Red Cross, BKV digital	Funding	USA	http://bkv.bkvdigital.com/blog/bkv-and-american-red-cross-to-present-using-analytics-to-mine-donor-data/
Predictive analytics in humanitarian supply chains[37]	Direct Relief and General Electric	Logistics	Global	https://www.gelifesciences.com/en/us/news-center/ai-in-healthcare-and-pandemics-10001
PRIO conflict prediction	Peace Research Institute Oslo (PRIO)	Conflict	Global	http://folk.uio.no/hahegre/Papers/PredictionISQ_Final.pdf
Project Jetson	UNHCR Innovation Service, UN Global Pulse, Essex University Big Data and Technology Centre, Omdena Foundation	Displacement and migration	Within and outside of Somalia	http://jetson.unhcr.org/
REACH flood susceptibility model	REACH, Shelter Cluster Yemen	Floods	Yemen	Link no longer available

[36] It is unclear who the work was for.

[37] Not the initiative's official name, no name provided by the organisation.

Sentry	Hala Systems, Inc.	Conflict	Syria	https://halasystems.com/
Social Media Monitoring - Quantifying Sentiment - Xenophobia in Europe /Crimson Hexagon	UNHCR' Innovation Service, UN Global Pulse, and Crimson Hexagon	Displacement and migration	Europe	https://www.unhcr.org/innovation/experiments/ https://www.unhcr.org/innovation/wp-content/uploads/2017/09/FINAL-White-Paper.pdf
The Early Warning Project	Simon-Skjodt Center and Dartmouth College	conflict	Global	https://earlywarningproject.ushmm.org/about
The inundation model - Google flood forecasting model	Google AI for Social Good, Indian Central water commission	Floods	Bihar, India	https://ai.googleblog.com/2019/09/an-inside-look-at-flood-forecasting.html
The Joint Analysis of Disaster Exposure (JADE)	OCHA (Asia-Pacific), WFP (Asia-Pacific), Pacific Disaster Centre	Disaster Risk Reduction	20 countries in Asia pacific	https://www.pdc.org/wp-content/uploads/PDC-WFP-UNOCHA-Partnership.pdf
The Managing Risk through Economic Development (M-RED) program's Drone Initiative	Mercy Corps	Disaster Risk Reduction	Nepal, Timor-Leste	https://medium.com/@tsterl20/the-ngos-eye-in-the-sky-d6ec520dd4b0

The Platform for Real-time Impact and Situation Monitoring (PRISM)/Vulnerability Analysis Monitoring Platform for the Impact of Regional Events (VAMPIRE) - PRISM/VAMPIRE	WFP, The Office of the President of the Republic of Indonesia, the Food and Agricultural Organization (FAO) and UN Global Pulse	Disaster Risk Reduction	Sri Lanka, Cambodia, Mongolia, Afghanistan	https://innovation.wfp.org/project/prism
Uber for Waste: Using predictive analytics to streamline waste collection in the camps[38]	UNICEF Jordan	WASH	Jordan's Za'atari camp	https://medium.com/@unicefjordan1/smart-refugee-camps-applying-the-best-of-iot-and-ict-for-better-camp-management-e35e619e7310
Using Predictive Analytics to Identify Children at High Risk of Defaulting From a Routine Immunization Program: Feasibility Study	Harvard Medical School Center for Global Health Delivery–Dubai, Interactive Research and Development	Disease outbreaks	LMICs	https://publichealth.jmir.org/2018/3/e63/
Violence Early-Warning System (ViEWS)	Uppsala University - Department of Peace and Conflict Research	Conflict	Global	https://www.pcr.uu.se/research/views/about-views/

[38] This is not the initiative's official name, no name provided by the organisation.

Water Monitoring in Turkana and Wajir	Oxfam GB, SenosrInsight, Element Blue[39]	WASH	Kenya	https://sensorinsight.io/wp-content/uploads/2016/07/Oxfam.pdf
Zika Map	UNICEF, Google	Disease outbreaks	Latin America	https://www.reuters.com/article/us-health-zika-alphabet/google-says-its-engineers-working-with-unicef-to-map-zika-idUSKCN0W50OR

[39] SensorInsight's parent company.