Impact-based humanitarian forecasting using machine learning for floods

A literature survey

Leonardo Marcuzzi[1]

Supervisor(s): Dr. Cynthia Liem[1]**, Marijn Roelvink**[1]

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 22, 2025

Name of the student: Leonardo Marcuzzi
Final project course: CSE3000 Research Project
Thesis committee: Dr. Cynthia Liem, Marijn Roelvink, Jing Sun

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Abstract

With the worsening of climate change, the complications brought on by floods every year create an increasing need for forecasting systems that humanitarian organizations can use to help populations in danger. This research presents a literature review of machine-learning models for impact-based flood forecasting, and compares them with existing humanitarian projects. The results examine the characteristics of the models surveyed, while the discussion focuses on understanding how these characteristics can define whether the machine learning models proposed can actually be translated to humanitarian settings. The main takeaways include the prevalent choice of deep learning and ensemble models, used to improve the adaptability of the models, the problems with data availability and data quality in different areas considered, and the difference between lead times, usability, and scalability of the models proposed in contrast with already used humanitarian projects. This study then highlights the importance of transparency and reproducibility of the survey by detailing the queries and databases used, ensuring accessibility of selected articles, and explaining the selection criteria and methodology. Ultimately, the review concludes with the key insights on the connection between academic prototypes and real-life humanitarian projects, as well as key areas for future research.

1 Introduction

Floods can cause devastating damage to cities, the environment, and animal populations, and climate change only worsens this situation [1]. In addition, floods cause huge long-term problems, such as contamination of water supplies or spread of diseases among populations accessing the waters or crops [2]. It is complicated and expensive for communities to counter a flood event once it starts happening, so predicting them has become increasingly important.

1.1 Research Question

The main question that the research tries to answer is the following:

Under what conditions can machine learning techniques be used effectively for impact-based flood humanitarian forecasting?

In order to produce a satisfactory answer and to carry out the literature review appropriately, multiple subquestions have emerged:

1. What types of machine learning models are used?
2. What are the sources and challenges of data collection?
3. How are impacts considered in these models?
4. What practical considerations influence the adoption of these models in humanitarian settings?

The first research question (RQ1) involved understanding the differences and types of machine learning models used throughout the different solutions. The second one (RQ2) focused on the sources, problems, and characteristics of the data collection pipeline. RQ3 involved the aspects of impact-based forecasting between different models and how each solution approaches them. Lastly, RQ4 considered the limitations of the models chosen and the comparison with already existing humanitarian projects for a better understanding of real-life usability.

1.2 Research Gap

Historically, predicting floods was done through the use of hydrological models, which represent the act of analyzing the water flow in a river or stream system using mathematical and computer tools [3]. This approach focuses directly on the hazards and what they may be, while recent research is moving toward what hazards may do and the warning systems around them [4]. In the past few years, it has been shown that machine learning algorithms have the ability to analyze large amounts of hydrological, meteorological, and topographical data, enabling solutions that improve upon accuracy and reliability while also quantifying the impacts [5]. This is the exact reason why more and more papers have been focusing on impact-based forecasting using machine learning in recent years, while, at the same time, comprehensive literature reviews on the topic are still lacking. For example, the recent systematic review on deep learning applications [6] for flood forecasting by Kumar et al. [7] provides an extensive review on solutions that use deep learning to predict floods or create mappings, being a great starting point to look into a big set of models, but it also does not focus on impact-based solutions and does not apply knowledge from gray literature.

The systematic review by Aljohani et al. [8] provides a comprehensive overview of flood prediction modeling techniques, categorizing them into hydrologic and machine learning models, and objectively assessing their advantages and disadvantages. Their work also explores the potential of hybrid strategies and includes a bibliometric analysis, offering valuable insights for researchers and practitioners in the field. Similarly, El Baida et al. [9] present a systematic literature review on classification machine learning models for urban flood hazard mapping, a novel contribution, as it systematically explores this specific area. This review is structured according to established guidelines and provides a detailed methodology including research questions and a robust search strategy across five major digital libraries, ultimately evaluating different ML classifiers and their performance. Bukhari et al. [10] offers a comprehensive data analytic perspective on flood monitoring and prevention, uniquely integrating discussions on various techniques such as IoT [11], machine learning, remote sensing, and early warning systems [12], thereby bridging gaps between different approaches and highlighting their utility in real-time data collection and informed decision-making.

Despite their valuable contributions, a recurring challenge, also evident in the work by Kumar et al. and in the review by Aljohani et al., is the difficulty in consistently reproducing their findings. The reporting of the search strategy for both of these reviews is often unclear and imprecise, which would make it complicated to recreate the same settings for a new review with similar objectives. Furthermore, while models discussed in these reviews are often intended for use by humanitarian agencies or local governments, many reviews on the topic, including El Baida et al. and Bukhari et al., do not incorporate insights from gray literature, which can be crucial for practical application and real-world scenarios.

This research focuses on conducting a literature review on impact-based forecasting using machine learning techniques for flood impacts. Ideally, this research would fill the gaps left by other papers, while also investigating them as a starting point for a systematic review on the matter. An important point that will be addressed is the actual feasibility of the methods surveyed, precisely in humanitarian settings, with the help of gray literature such as journals, conferences, or other reports from organizations.

2 Methodology

This study is based on a literature review conducted using the SALSA framework [13]. The following subsections highlight the different parts of the SALSA methodology and their respective descriptions related to the research itself. A flowchart of the whole process is displayed under this section (see Figure 1).

2.1 Search

The Search part of the SALSA methodology incorporates all activities related to the scouting and selection process of relevant papers for inclusion in the succeeding stages of the review. This involves the queries used and the search strategy.

2.1.1 Queries

The queries chosen for the search were run on Scopus because it was shown to produce the most relevant results. The first query was aimed at finding most of the papers relevant to the exact purpose of the research:

("flood") AND ("impact-based") AND ("forecasting" OR "machine learning" OR "AI" OR "big data") AND ("humanitarian" OR "aid") AND PUBYEAR > 2018 AND PUBYEAR < 2026

This query yielded 166 results. Then, another query was created to find the papers that mentioned impact-based flood forecasting, but did not mention humanitarian aid. Additionally, synonyms of machine learning used in the previous query were removed to avoid having the results inflated from other types of forecasting. Ultimately, this was the second query chosen to compensate the deficiencies of the first one:

("flood") AND ("impact-based") AND ("machine learning") AND ("early warning") AND PUBYEAR > 2018 AND PUBYEAR < 2026

135 papers were found through this query, amounting to a total of 301 articles before any of the eligibility criteria were applied.

2.1.2 Search Strategy

After having run the queries, the search strategy involved examining titles and abstracts of the papers retrieved to select an initial subset of articles. Then, out of this subset, they were evaluated to determine compliance with the inclusion and exclusion criteria. Moreover, the papers were briefly read to understand whether they were relevant to the subquestions or to the literature review in general. This last subset was then used in the appraisal phase.

The inclusion criteria of the search were strictly related to the queries created and to the way papers were then selected from the search results. These were the inclusion criteria chosen for this research:

- Papers focusing on machine learning techniques for flood forecasting.

  OR

- Papers related to humanitarian applications of flood forecasting.

  OR

- Papers mentioning the limits of humanitarian adoption of flood forecasting.

OR

- Papers mentioning impact-based flood forecasting.

  OR

- Publications from humanitarian sources about flood forecasting.

And then, the criteria that each paper was required to satisfy:

- Papers published from 2019 onward.

  AND

- Papers written in English.

Subsequently, the exclusion criteria were applied to the papers selected through the inclusion process. These were the exclusion criteria for this research:

- Duplicates across different databases.

  AND

- Papers that are only summaries of others.

The search strategy resulted in an initial subset of 35 papers to be analyzed in the successive phases of SALSA. Additionally, 24 articles coming from gray literature sources were found manually from the Directory of AI-Enabled Humanitarian Projects, the Catalogue of Predictive Models in the Humanitarian Sector, and the Anticipation Hub. These reports still adhered to inclusion and exclusion criteria, but no automated query was used. Nevertheless, the keywords from the queries mentioned before were used in these different databases to obtain the 24 new results.

2.2 Appraisal

In the appraisal phase, the initial subset of papers was reviewed briefly by reading the important sections, such as the abstract, introduction, and results, taking notes of the useful elements, and saving notable references. The snowballing of references during the appraisal phase added 12 papers to the total number reviewed. Additionally, the remaining papers were ordered in different categories based on the elements covered:

- Papers describing single machine learning forecasting models, their use, and limitations.
- Papers that perform systematic literature reviews on models.
- Reports from gray literature sources that mention use cases of models.

2.3 Synthesis

In the synthesis phase, the main findings of the papers retained after appraisal were systematically organized to prepare for a structured analysis. Each selected article and report was read in full, and its key insights were extracted using a predefined annotation scheme based on the research questions. Specifically, the synthesis involved:

- Extracting information relevant to the subquestions, such as technical descriptions of machine learning (ML) models, operational challenges, and humanitarian applications.

- Summarizing the key contributions, limitations, and context of use for each source.
- Identifying and tagging elements of the text that aligned with specific research questions.
- Compiling a draft comparative table to track which research questions each paper contributed to, and to facilitate clustering by theme.

The outcome of the synthesis phase was a refined subset of 10 academic articles and 9 gray literature reports, each clearly linked to specific research questions. This laid the foundation for the more in-depth analysis that followed.

2.4 Analysis

When analyzing the final subset of papers, the main takeaways considered were models used, type of data sources, the geographical location of the use case, the type of impact considered, the eventual limitations of the solution, and whether the solution is ready for humanitarian use. These findings were useful to compile a table (see Table 1) highlighting the most important papers in the review and to provide a useful overview to answer the research questions.

From this annotated and grouped information, results were drawn out in direct response to the initial research questions, completing the final analysis and the research itself.

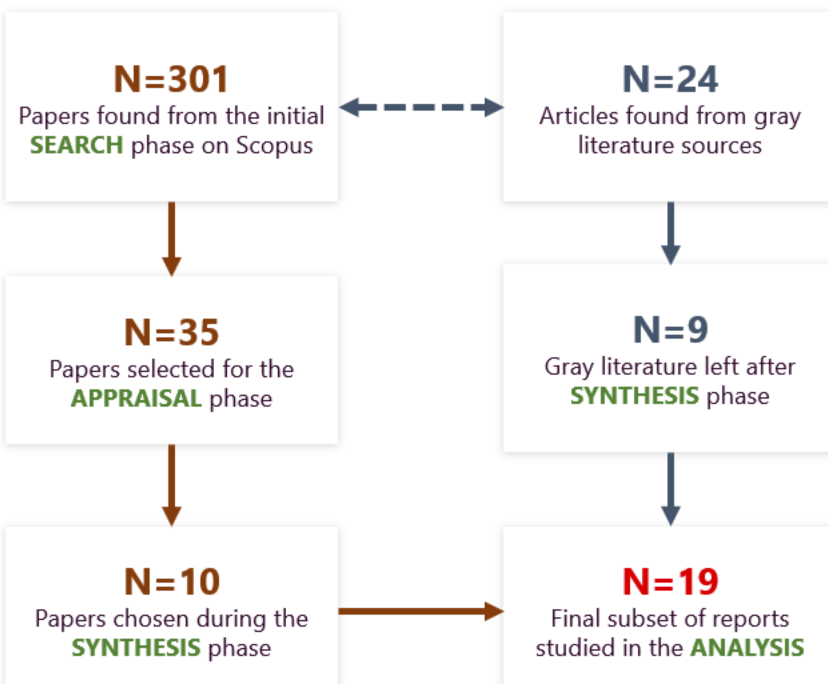

Figure 1: Flowchart illustrating the screening and selection process of the papers

3 Results

Starting with RQ1 and the models used, the key outcome is that most recent solutions adopt ensemble methods to combine multiple machine learning classifiers to forecast new data. The ones that do not, still tend to use bigger models like Convolutional Neural Networks [14] or Multi-Layer Perceptrons [15]. This is because ensemble learning significantly improves predictive performance by combining the strengths of multiple base models, reducing overfitting, and enhancing robustness across various tasks. Similarly, deep learning models such as CNNs and MLPs can automatically extract high-level features from complex data [16], making them well-suited for tasks like flood forecasting. However, both approaches have trade-offs as mentioned by the report of Mohammed and Kora [16]: while ensemble methods improve generalization by aggregating diverse learners, deep learning models often require careful hyperparameter tuning and large computational resources [16].

Moving on to RQ2, the data sources used for flood forecasting vary widely, including satellite imagery, rainfall records, flood maps, and historical data. This diversity demonstrates that effective forecasting can rely on multiple data types, which is a promising insight given that data scarcity remains a major barrier to humanitarian adoption, particularly in regions lacking real-time or high-quality information. Flexibility in these use cases is the key to achieving positive results in different conditions and regions. In order to overcome the problems with data availability, the solution by Roy et al [17] does not incorporate all of the features selected (with drainage data), or adjusts others to still provide a positive input (with rainfall data).

When looking at RQ3, not all papers explicitly adopt an impact-based forecasting framework, but many incorporate impact-related variables to tune or feed their models. In this context, impacts are defined as the measurable consequences of flooding on natural systems, infrastructure, and human activity, including changes in water level, inundation extent, damage to transport networks, displacement of people, and economic loss. One exception is the study by Noor et al. [18], which primarily focuses on predicting water levels using hydrological data, without addressing downstream consequences of flooding. Despite this, their approach demonstrates potential utility in humanitarian applications, where timely water level forecasts can inform evacuation or relief decisions. In contrast, the majority of the reviewed papers address physical and hydrological impacts, including variables such as river discharge, rainfall intensity, flood extent, and inundation mapping, elements that influence or directly represent the physical dynamics of flood events. Some studies go further by quantifying infrastructure-specific impacts, such as disruptions to transportation networks [17], or even socioeconomic consequences, such as economic losses, population exposure, or vulnerability metrics [19].

Lastly, mentioning RQ4, the main limitations of the solutions mentioned from the authors are the data quality and availability, which are a big problem in areas without a huge database of hydrological information. Approaches like the one from Nahak et al. [20] or de Lima et al. [15] use data sources from smaller areas that do have better coverage, thanks to humanitarian or governmental agencies, but would fail to be transposed in other regions without this rich availability. Another important limitation is the scalability of the solution proposed, as shown in the work by El Baida et al [14]. Additionally, the paper by Ren et al. [21], describes problems with the biases of sampling the data.

	Models	Data Source	Case Study	Impact Type	Limitations
Lee et al. [19]	LSTM [22]	Geospatial, historical	Dorim, South Korea	Hydrological, Socioeconomic	Data quality, complexity
Ren et al. [21]	RF[23], XGBoost [24]	Satellite imagery	Kunming, China	Hydrological	Data availability
Mia et al. [25]	ADT [26], NB [27], ANN [28], DLNN [29]	Geospatial	Padma, Bangladesh	Hydrological	Data availability
Roy et al. [17]	LSTM, seq2seq [30]	Sequential, runoff	Norfolk, Virginia	Transportation	Data Availability
El Baida et al. [14]	CNN [31]	Rainfall, flood maps	Zaio, Morocco	Hydrological	Data quality, scalability
de Lima et al [15]	MLP [32]	Water level, rainfall	Brazil	Hydrological	Data availability
Noor et al. [18]	STALSTM [33]	Water Level	Bangladesh	N/A	Data availability
Won et al. [34]	ANN, LSTM	Hydrological, Geospatial	Dorim, South Korea	Hydrological	Data quality
Mangukiya et al. [35]	RF, DT [36], XGBoost	Geospatial	Surat, India	Hydrological	Data quality
Nahak et al. [20]	FFNN [37], CNN	Regional, hydrological	N/A	Hydrological	Data availability

Table 1: Overview of flood forecasting studies analyzed in the review

4 Discussion

One of the most important goals of this research is to understand the conditions that make a machine learning model usable in humanitarian settings. To understand the connection between the models shown in the papers analyzed and humanitarian action, it is necessary to compare the two using real-life examples and to use the following criteria:

1. Direct mention or explanation of the model used for humanitarian use in the article.
2. Scalability of the solution proposed.
3. Lead time of a possible trigger warning.
4. Possible fast deployment of the solution proposed.
5. Comparison of the limitations shown by a model with humanitarian examples.
6. Comparison of the type of model and data used with real-life humanitarian examples.

The first five points are directly addressed in the table below this subsection (see Table 2), while comparison and readiness of models are discussed in later paragraphs. The first criterion is fulfilled in a paper if it directly mentions the focus for humanitarian help, and presents element that suggest an integration with real-time aid. The second one, scalability, is used to describe whether a solution proposed could actually be transposed to other cases, and it is not location or data specific, as described directly in the article. Lead times are not present in all papers, but they are useful for comparison with humanitarian projects. Moreover, fast deployment is a concept related to multiple aspects described in the articles like complexity, data sources, and availability, which describe the speed of integration in a real situation of a model. Lastly, limitations consider additional aspects that limit the use of a model in humanitarian settings.

One example of the first criterion applied to a paper is the work of Lee et al. [19], which provides a full section showing the applications of their model, directly mentioning early warning for humanitarian action with lead times from 10 to 90 minutes. There are data and complexity limitations for this model, but in the case studies it still performed positively. This paper is a great case of a well-tested model, especially for further humanitarian use, and it states that in the results too, so it is marked as aid-ready.

Other models like the one by Ren et al. [21] and by Mia et al. [25] do not directly mention humanitarian use in their papers. Ren et al. model requires lots of data and resources, has issues with the sampling of the dataset, and accuracy is reduced with incomplete historical data, while the Mia et al. model needs complex hydrodynamic data, shows a bias in the non-flood location selection, and its deep learning models do not adapt to different locations. On the other hand, already deployed solutions like UNOSAT and HYDRAFloods have clear humanitarian focus, a fast deployment, higher data availability, minimal calibration, automatic activation for humanitarian response, and useful web dashboards. Ren et al. and Mia et al. are not yet ready for humanitarian use following the three criteria mentioned above, while they still represent useful case studies for lab work and academia.

On the other hand, the model of Roy et al. [17] focuses on street-scale nuisance flood forecasting using impacts on transportation, providing an improved connection with humanitarian action. For a small subset of 22 streets, the Roy et al. solution raises warnings for a storm event from 0.09 to 0.11 seconds for the short-term forecast to 0.30 to 0.35 seconds for the long-term one, proving its low computational complexity. On the other hand, in the IFRC reports for Early Action Protocols for flood forecasting in Djibouti [38] and in Bangladesh [39] the deployed solutions have respectively a 7-day and a 5-day lead, emphasizing the need for enough time to organize humanitarian help. Still, the report from Roy et al. provides warnings that can be used in real-time flood problems, which can be helpful in different scenarios. Nevertheless, to integrate its solution in humanitarian settings, much more data needs to be fed to the model, as well as a full integration with an operating agency and proper disclosure of the data collected. Even with these limitations, this solution proves to be ready in a city-like setting for humanitarian use, so it is marked accordingly.

Other solutions like El Baida et al. [14] and de Lima et al. [15] are partially ready to be used in humanitarian settings, but still lack decisive characteristics explained before. The first one has the ability to predict both pluvial and fluvial floods, but has the problem of being too location-specific, of needing a complex training, and lots of data. Moreover, the second one has a similar problem with the ability to adapt to different locations, and with the fact that it works on a 1 to 3-hour lead, while EAP in Bangladesh, Indonesia [40], and Djibouti show, respectively, a 5, 6, and 7-day lead.

Other papers show the possibility of being used as a support for humanitarian operations, but they are not ready on their own. For example, the model by Noor et al. [18] does not consider impacts, but only predicts the water level, which could still contribute to early

warning systems for floods, but that does not yet fulfill the criteria mentioned before. On the other hand, the model by Nahak et al. [20] is one of the latest papers considering federated learning for flood forecasting, which combines smaller models from multiple sources, trains them with regional data, and issues flood alerts with a 5-day lead time. This appears as a great solution, but that has yet to find a real-life case study to be applied and tested, so it is not being marked as aid-ready.
Lastly, two papers show promising and possibly working results in humanitarian settings. Starting with the work by Won et al. [34], which has a process-oriented approach, from the data collection to the warning system, integrating real-time information and creating inundation maps. In order to be used in humanitarian settings, it is only missing a proper integration with humanitarian agencies or with an alert and warning system that is more interactive, as well as a more refined data collection pipeline, but the process would be possible. To conclude, the work by Mangukiya et al. [35] is showing directly actionable input, producing flood inundation maps with depth information, estimating the affected populations, considering evacuations, and planning logistical routes. It is a system of fast deployment, and can be scaled to multiple situations. On the other hand, it is also missing a proper real-time integration and suffers from a data quality problem, but projects like Google's Flood Forecasting pipeline, which also provides inundation maps, show a similar solution already deployed with real-time alerts, proving that a humanitarian implementation is indeed possible.

	Aid-focus	**Scalability**	**Lead time**	**Fast Deployment**	**Limitations**	**Aid-ready**
Lee et al. [19]	✓	✓	10-90 minutes	✓	Integration, short warning	✓
Ren et al. [21]	X	X	N/A	X	Humanitarian focus	X
Mia et al. [25]	X	X	N/A	X	Humanitarian focus	X
Roy et al. [17]	✓	✓	0.09-0.35 seconds	✓	Short warning	✓
El Baida et al. [14]	X	X	N/A	X	Humanitarian focus	X
de Lima et al [15]	X	X	1-3 hours	X	Integration, short warning	X
Noor et al. [18]	✓	✓	N/A	✓	Not redy on its own	X
Won et al. [34]	✓	✓	30 minutes	✓	Integration	✓
Mangukiya et al. [35]	✓	✓	N/A	✓	Integration	✓
Nahak et al. [20]	✓	X	5-7 day	X	No case study	X

Table 2: Overv iew of the aid-readiness of flood forecasting models studied

5 Responsible Research

The first aspect to consider when conducting a literature review is the reproducibility of the findings. To support reproducibility, this paper ensures transparency in the queries and databases used, all of which are documented. Additionally, the articles selected are open-access or accessible through institutional subscriptions, and some were retrieved through references within initially selected papers rather than direct query results. The only exceptions to this last statement are the three articles from reports of the Red Cross that directly come from the authors. Although they are not present in an online version right now, it is still possible to request them from the related agencies.
Another related but distinct aspect is the transparency of the research process itself. This includes not only detailing the limitations and challenges of the models surveyed but also being explicit about the selection criteria and methodology followed in this review. Still, this is not intended to be a fully systematized literature review, so it might lose precision, and biases of the author are part of the process to select the articles.
Many of the chosen papers support transparency by providing access to their code and data, allowing for independent verification of their findings. The papers in this review are considered trustworthy, as well as the data sources that they provide in their research. Additionally, the data used in this review does not influence or is in any way related to individuals.
Of course, there risks related to the review itself, as it is considering aspects of humanitarian help, which involve disasters and human lives. On the other hand, every research in this field is essential, as every new insight contributes to improving humanitarian forecasting and, ultimately, the effectiveness and timeliness of aid delivery.
Lastly, this paper does not use in any way the help of LLMs, not to rewrite, not to generate ideas, or not to assist in the analysis process in any way.

6 Conclusions and Future Work

This research focused on identifying the conditions necessary for the effective adoption of impact-based flood forecasting using machine learning in humanitarian contexts. In order to understand this problem, it was necessary to review the papers that, in the past few years, have developed models that would show promising results in humanitarian settings. Secondly, multiple sources of gray literature were reviewed to grasp the difference between projects that are already deployed and others that are only prototypes in academia.
Answering RQ1, most of the models used by the articles reviewed were ensemble and deep learning models, as they were shown to be able to handle multiple sources of data simultaneously and adapt to different situations and locations, which is fundamental in humanitarian settings.
To cite RQ2, a recurring problem in the papers surveyed was data collection, as availability and the quality of this information were shown to be huge problems in certain regions of the world. There was either no historical collection of flood data, or it was simply hard to retrieve qualitative data for multiple regions, so the case studies were usually located in smaller and controlled environments, which could avoid some of the data problems. Of course, this also meant that these models could not scale properly in humanitarian settings.
Another important point of the research was the consideration of impacts in the forecasting pipeline, as mentioned in RQ3, which has been used more and more in the past few years, as previously the focus was primarily on hazards. Most of the models surveyed did consider an impact-based approach that favors the creation of warning systems, as it is necessary

for humanitarian agencies to understand the possible damages that floods could cause in a specific region to organize the response on time. The main impacts considered by the papers were hydrological, so directly on the physical damages that the flood would cause, but others, totally or partially, also considered socioeconomic and transportation impacts.

Finally, answering RQ4, the main focus of the discussion was on the practical considerations that limit the use of the models reviewed in actual humanitarian settings. Most of the articles did not consider a proper integration with humanitarian agencies, making a possible integration for a real-life use quite harder. The ones who did, on the other hand, usually described the whole pipeline from the data collection to the warning system in the same article, making it easier to adopt for humanitarian organizations. Another issue was that no model showed graphical interfaces to test their solution, which is a very common practice for already deployed projects, as well as the ability to receive multiple different sources of data and locations as input. Additionally, some big differences between the articles surveyed and humanitarian projects were the lead times; most of the newer models were focused on alerting to flood events minutes or hours before they happen, while most of the accepted and used solutions preferred multiple-day warnings. This last point proves the preference of the models that give warnings over a longer period, since they allow humanitarian organizations to better manage their forces and finances.

It is important to mention that, out of the 10 articles reviewed, only 4 of them are marked as ready to be used in humanitarian settings, and they would still need proper integration in a real-life environment. Most of the papers surveyed still show a case study on a specific location, but most of the humanitarian projects already deployed are working in multiple different regions. This happens because it is necessary to be efficient with the resources and organization, especially in a time of huge new budget cuts [41]. This is exactly why recently updated humanitarian projects like InaSAFE provide ways of combining data from multiple sources, as well as creating natural hazards impact scenarios, creating a whole pipeline that is ready to use. Of course, it is still incredibly important to research new machine learning models that can provide interesting results even in localized settings, but papers that do provide space for limitations, integration, data collection, and scalability are going to be much more easily used in real-life settings.

Another useful approach used in humanitarian contexts is combining multiple databases to gain information about different early warning systems at the same time. For example, the Anticipation Hub shows the lead time of multiple EAP triggers, which is incredibly useful when trying to facilitate the exchange in the development process by practitioners. Another significant project that aggregates information is WorldFloods, a newly compiled dataset of 119 globally verified flooding events from disaster response organizations [42]. A similar type of collaboration would also be useful using the models developed by university groups, as it can lead to new, promising developments.

To conclude, this research only covers a few years of advances on impact-based humanitarian forecasting using machine learning, and the amount of material found was still very significant, so new reviews could be written in the next few years to find new conclusions on the topic. Additional surveys could cover the characteristics and metrics used by machine learning models to understand further what aspects are favored in the field of humanitarian forecasting and which are not, as well as the pipeline to bring a machine learning model to be used in real-life settings.

References

[1] R. L. Wilby and R. Keenan, "Adapting to flood risk under climate change", *Progress in Physical Geography*, 2012. DOI: 10.1177/0309133312438908.

[2] C. J. Talbot, E. M. Bennett, K. Cassell, D. M. Hanes, E. C. Minor, H. Paerl, P. A. Raymond, R. Vargas, P. G. Vidon, W. Wollheim, and M. A. Xenopoulos, "The impact of flooding on aquatic ecosystem services", *Biogeochemistry*, 2018. DOI: 10.1007/s10533-018-0449-7.

[3] P. A. Herrera, M. A. Marazuela, and T. Hofmann, "Parameter estimation and uncertainty analysis in hydrological modeling", *WIREs WATER*, 2021. DOI: 10.1002/wat2.1569.

[4] S. H. Potter, T. Kox, B. Mills, A. Taylor, J. Robbins, C. Cerrudo, F. Wyatt, S. Harrison, B. Golding, W. Lang, A. J. Harris, R. Kaltenberger, S. Kienberger, H. Brooks, and A. Tupper, "Research gaps and challenges for impact-based forecasts and warnings: Results of international workshops for high impact weather in 2022", *International Journal of Disaster Risk Reduction*, 2025. DOI: 10.1016/j.ijdrr.2025.105234.

[5] V. Kumar, K. V. Sharma, T. Caloiero, D. J. Mehta, and K. Singh, "Comprehensive overview of flood modeling approaches: A review of recent advances", *Hydrology*, 2023. DOI: 10.3390/hydrology10070141.

[6] S. Razavi, "Deep learning, explained: Fundamentals, explainability, and bridgeability to process-based modelling", *Environmental Modelling Software*, 2021. DOI: 10.1016/j.envsoft.2021.105159.

[7] V. Kumar, H. M. Azamathulla, K. V. Sharma, D. J. Mehta, and K. T. Maharaj, "The state of the art in deep learning applications, challenges, and future prospects: A comprehensive review of flood forecasting and management", *Sustainability*, 2023. DOI: 10.3390/su151310543.

[8] F. H. Aljohani, A. B. Alkhodre, A. A. A. Sen, M. S. Ramazan, B. Alzahrani, and M. S. Siddiqui, "Flood prediction using hydrologic and ml-based modeling: A systematic review", *International Journal of Advanced Computer Science and Applications*, 2023. DOI: 10.14569/IJACSA.2023.0141155.

[9] M. E. Baida, M. Hosni, F. Boushaba, and M. Chourak, "A systematic literature review on classification machine learning for urban flood hazard mapping", *European Water Resources Association*, 2024. DOI: 10.1007/s11269-024-03940-7.

[10] S. A. S. Bukhari1, I. Shafi, J. Ahmad, S. G. Villar, E. G. Villena, T. Khurshaid, and I. Ashraf, "Review of flood monitoring and prevention approaches: A data analytic perspective", *Natural Hazards*, 2024. DOI: 10.1007/s11069-024-07050-w.

[11] S. Kumar, P. Tiwari, and M. Zymbler, "Internet of things is a revolutionary approach for future technology enhancement: A review", *Journal of Big Data*, 2019. DOI: 10.1186/s40537-019-0268-2.

[12] M. M. Islam, M. Hasan, M. S. Mia, A. A. Masud, and A. R. M. T. Islam, "Early warning systems in climate risk management: Roles and implementations in eradicating barriers and overcoming challenges", *Natural Hazards Research*, 2025. DOI: 10.1016/j.nhres.2025.01.007.

[13] M. J. Grant and A. Booth, "A typology of reviews: An analysis of 14 review types and associated methodologies", *Health Information & Libraries Journal*, 2009. DOI: 10.1111/j.1471-1842.2009.00848.x.

[14] M. E. Baida, F. Boushaba, M. Chourak, and M. Hosni, "Real-time urban flood depth mapping: Convolutional neural networks for pluvial and fluvial flood emulation", *Water Resources Management*, 2024. DOI: 10.1007/s11269-024-03886-w.

[15] G. R. T. de Lima, R. de Oliveira Caran, and L. F. de Aguiar Filho, "Use of neural network as a support tool in water level forecasting and issuing flash floods early warnings to three small brazilian urban watersheds", *Earth Science Informatics*, 2023. DOI: 10.1007/s12145-023-01159-5.

[16] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges", *Journal of King Saud University - Computer and Information Sciences*, 2023. DOI: 10.1016/j.jksuci.2023.01.014.

[17] B. Roy, J. L. Goodall, D. McSpadden, S. Goldenberg, and M. Schram, "Forecasting multi-step-ahead street-scale nuisance flooding using a seq2seq lstm surrogate model for real-time application in a coastal-urban city", *Hydrology*, 2025. DOI: 10.1016/j.jhydrol.2025.132697.

[18] F. Noor, S. Haq, M. Rakib, T. Ahmed, Z. Jamal, Z. S. Siam, R. T. Hasan, M. S. G. Adnan, A. Dewan, and R. M. Rahman, "Water level forecasting using spatiotemporal attention-based long short-term memory network", *Water*, 2022. DOI: 10.3390/w14040612.

[19] J. H. Lee, G. M. Yuk, H. T. Moon, and Y.-I. Moon, "Integrated flood forecasting and warning system against flash rainfall in the small-scaled urban stream", *Atmosphere*, 2020. DOI: 10.3390/atmos11090971.

[20] S. K. Nahak, S. K. Acharya, and D. Padhy, "Enhanced flood forecasting: Revolutionizing prediction with federated learning", *Smart Trends in Computing and Communications*, 2024. DOI: 10.1007/978-981-97-1323-3_39.

[21] H. Ren, B. Pang, P. Bai, G. Zhao, S. Liu, Yuanyuan, and M. Li, "Flood susceptibility assessment with random sampling strategy in ensemble learning (rf and xgboost)", *Remote Sensing*, 2024. DOI: 10.3390/rs16020320.

[22] R. C. Staudemeyer and E. R. Morris, "Understanding lstm – a tutorial into long short-term memory recurrent neural networks", *arXiv*, 2019. DOI: 10.48550/arXiv.1909.09586.

[23] L. Breiman, "Random forests", *Machine Learning*, 2001. DOI: 10.1023/A:1010933404324.

[24] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system", *arXiv*, 2016. DOI: 10.48550/arXiv.1603.02754.

[25] U. Mia, T. N. Chowdhury, R. Chakrabortty, S. C. Pal, M. K. Al-Sadoon, R. Costache, and A. R. T. Islam, "Flood susceptibility modeling using an advanced deep learning-based iterative classifier optimizer", *Land*, 2023. DOI: 10.3390/land12040810.

[26] M. P.-L. Ooi, H. K. Sok, Y. C. Kuang, and S. Demidenko, "Chapter 19 - alternating decision trees", *Handbook of Neural Computation*, 2017. DOI: 10.1016/B978-0-12-811318-9.00019-3.

[27] O. Peretz, M. Koren, and O. Koren, "Naive bayes classifier â an ensemble procedure for recall and precision enrichment", *Engineering Applications of Artificial Intelligence*, 2024. DOI: 10.1016/j.engappai.2024.108972.

[28] R. Qamar and B. A. Zardari, "Artificial neural networks: An overview", *Mesopotamian Journal of Computer Science*, 2023. DOI: 10.58496/MJCSC/2023/015.

[29] D. T. Bui, N.-D. Hoang, F. Martinez-Alvarez, P.-T. T. Ngo, P. V. Hoa, T. D. Pham, P. Samui, and R. Costache, "A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high-frequency tropical storm area", *Science of The Total Environment*, 2020. DOI: 10.58496/MJCSC/2023/015.

[30] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks", *arXiv*, 2014. DOI: 10.48550/arXiv.1409.3215.

[31] K. O'Shea and R. Nash, "An introduction to convolutional neural networks", *arXiv*, 2015. DOI: 10.48550/arXiv.1511.08458.

[32] M. Gardner and S. Dorling, "Artificial neural networks (the multilayer perceptron)âa review of applications in the atmospheric sciences", *Atmospheric Environment*, 1998. DOI: 10.1016/S1352-2310(97)00447-0.

[33] Y. Ding, Y. Zhu, J. Feng, P. Zhang, and Z. Cheng, "Interpretable spatio-temporal attention lstm model for flood forecasting", *Neurocomputing*, 2020. DOI: 10.1016/j.neucom.2020.04.110.

[34] Y.-M. Won, J.-H. Lee, H.-T. Moon, and Y.-I. Moon, "Development and application of an urban flood forecasting and warning process to reduce urban flood damage: A case study of dorim river basin, seoul", *Water*, 2022. DOI: 10.3390/w14020187.

[35] N. K. Mangukiya, S. Kushwaha, and A. Sharma, "A novel multi-model ensemble framework for fluvial flood inundation mapping", *Environmental Modelling and Software*, 2024. DOI: 10.1016/j.envsoft.2024.106163.

[36] J. R. Quinlan, "Induction of decision trees", *Machine Learning*, 1986. DOI: 10.1007/BF00116251.

[37] M. H. Sazli, "A brief review of feed-forward neural networks", *Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering*, 2006. DOI: 10.1501/commua1-2_0000000026.

[38] IFRC, "Simplified early action protocol, djibouti", Unpublished internal report, requested from author, 2022.

[39] IFRC, "Forecast-based financing/action (fbf/a) / flood early action protocol (eap), bangladesh", Unpublished internal report, requested from author, 2021.

[40] Indonesian-Red-Cross, "Anticipatory action, early action protocol, floods, indonesia", Unpublished internal report, requested from author, 2024.

[41] J. Rankin, "Eu will struggle to fill gap left by usaid as european countries cut their budgets", *The Guardian*, Apr. 15, 2025. [Online]. Available: https://www.theguardian.com/global-development/2025/apr/15/eu-struggle-fill-gap-usaid-european-countries-cut-budgets.

[42] G. Mateo-Garcia, J. Veitch-Michaelis, L. Smith, S. V. Oprea, G. Schumann, Y. Gal, A. G. Baydin, and D. Backes, "Towards global flood mapping onboard low cost satellites with machine learning", *Nature*, 2021. DOI: 10.1038/s41598-021-86650-z.