*The OCHA Centre for Humanitarian Data’s* [*Peer Review Framework for Predictive Analytics in Humanitarian Response*](https://centre.humdata.org/peer-review-framework-2020) *aims to create standards and processes for the use of models in our sector. The Framework consists of six steps: model submission, technical review, implementation plan submission, ethical review, client consultation, and final report. The output of the model submission is a completed Model Card that includes information about the intended use, model development, model evaluation, and operational readiness. For further information, please contact Leonardo Milano:* [*leonardo.milano@un.org*](mailto:leonardo.milano@un.org)*.*

| **510 Typhoon Impact Model | October 2021** |
| --- |
| Model Summary   * Person/organization developing the model: Aklilu Teklesadik, Netherlands Red Cross - 510 Global * Model date: Last updated: 10 Sep 2021 * Model version: v1.0 * Model type: Machine learning (XGBoost Tree) * License: GPL v3 * Main contacts: Aklilu Teklesadik * Link (if available): [GitHub repository](https://github.com/rodekruis/Typhoon-Impact-based-forecasting-model)   1. Intended Use   1. In-scope use cases: what is the actual and potential scope of the model? Describe the situations in which the model output is expected to be reliable.   The 510 Philippines Typhoon Impact Model is designed to predict the per-municipality percentage of totally damaged houses based on a given typhoon track. The input parameters are both meteorological (rainfall, windspeed) and structural (housing material, number of houses). It uses the ECMWF ensemble typhoon track forecasts up to 72 h before landfall to provide probabilistic estimates of the percentage of houses damaged per municipality and in total.  The model output will be reliable for typhoon events in the category 2- 4 range, as the training datasets are lacking entries for events in higher and lower extreme ends. Additionally, the impact predicted by the model mainly concerns wind speeds, and the model demonstrates poor performance for events where the majority of the damage is not due to wind. For example, for typhoon Hayan the main cause of damage was storm surge, and in this case the typhoon model exhibited high mean absolute error.   1. Out-of-scope use cases: what are the model’s limits and constraints?   Describe the situations in which the model output may not be reliable?  The model is trained to predict damage per municipality as damage loss data for historical events was available only at a municipality level, and may not perform as well for aggregates. This model is limited to the Philippines.  The model performance may be poor in predicting damage for tropical storms with low wind speed / high rainfall, as the damage/loss data is available mainly for higher category typhoons.  The model is based on a dataset of only 30 typhoon records and is limited in accuracy for that reason.  The model results depend heavily on the forecasted typhoon tracks, which are particularly unreliable for lead times above 72 h. Typhoons are rapidly-changing weather systems, and the forecast is only updated every 6 hours. Even then, it takes several hours for the forecast to run, so the initial conditions are often already outdated.   1. Model interpretation: what does the output represent?   The model outputs the probability that the number of totally damaged houses will exceed a certain number, based on the ensemble typhoon track forecast. Different probability / number of houses totally damaged threshold combinations are used as the trigger, which is meant to capture the number and certainty of people affected.  2. Model Development   1. Details of the datasets used to build the model.    1. Describe the sources of data, size and scope of the datasets.   The model is trained on over 40 different features at the municipality level. These include:  ***Pre disaster indicators***  Table 1 describes the data sets used to develop the typhoon impact model. These data sets were collected from different sources through desk research and in-country visits of key stakeholders. It is essential that these datasets have national spatial coverage at the same aggregation level. As not all data sets - in particular historical damage counts - are available at barangay level, which is the lowest administrative unit in the Philippines, the typhoon model can only be developed at the next administrative level, which is a municipality level.  ***Damage/ loss data***  Damage/loss data for historical typhoon events comes from the National DRR and Management Council (NDRRCM). These data sets have information on completely and partly damaged houses, as well as people affected. Analysis of these datasets indicates for some of the events, e.g. Typhoon Haiyan, the reported number of people affected are highly coarse estimates; where either 0%, 50% or 100% of the total population in the area are reported as being affected. For a learning algorithm this distribution cannot be used to provide a reliable prediction.  ***Typhoon hazard data***  For each of the historical typhoon events, several hazard variables such as windspeed, rainfall, duration of exposure, distance from first impact, and distance from typhoon eye were used to describe the level of exposure of each municipality.   * Wind speed and distance from typhoon track were calculated based on the International Best Track Archive for Climate Stewardship (IBTrACS) datasets, which contains time series data on the location and intensity of the typhoons. We used a parametric wind field model from Holland 2008 to extract information for each of the municipalities. * Rainfall data from the Global precipitation measurement (GPM) mission was used to extract intensity and volume of rainfall recorded in each of the municipalities during the typhoon events. We calculated the per-municipality maximum intensity over a period of 6 hours, 24 hours, and cumulative over the typhoon event.   ***Data on landslides / storm surges***  In addition to wind speed, landslides and flooding also contribute to damage during a typhoon event. In the model we used additional datasets to capture the potential damage caused by these agents by defining variables which indicate vulnerability of each municipality to landslide/storm surge related risks. This was done by calculating the percentage of houses located in landslide and storm surge risk zones. These variables were computed by combining hazard maps for landslide/storm surge from National Operational Assessment of Hazard (NOAH) and settlement layer data of Facebook.  ***Data for the operational model***  For the operational typhoon EAP trigger model, the hazard data( typhoon track and rainfall) are obtained from forecasts, while pre-disaster indicators are the same as for the model training. The typhoon forecast data comes from numerical weather prediction models. There are multiple tropical cyclone forecasting systems which provide forecast information for typhoons. To select the best product, we conducted a skill assessment of 10 numerical weather prediction models from different forecasting agencies, which are all mandated by the world meteorological agency (WMO) to issue a forecast. This skill assessment was done by comparing archived forecasts of the different products against observation. Based on the skill assessment, the ECMWF forecast was selected for forcing the operational typhoon trigger model.  **Table 1:** Data sources   | **Data** | **Description** | **Temporal scale** | **Spatial Scale** | **Source** | | --- | --- | --- | --- | --- | | Wind/track data for historical Typhoon events | Dataset to describe hazard related variables for historical typhoons. I.e. windspeed, typhoon track | Every 6hrs (1951-2019) | Point coordinate for typhoon center | RSMC Best Track Data from Japan meteorological agency (1951-2019)<https://www.jma.go.jp/jma/jma-eng/jma-center/rsmc-hp-pub-eg/besttrack.html> | | Rainfall data during historical typhoon events | Maximum and intensity(per day) and total volume of rain during typhoon events | Daily (200o-to present) | 0.1° | Global Precipitation Measurement (GPM)  <https://pmm.nasa.gov/data-access/downloads/gpm> | | Common operational datasets | Admin boundaries with pcode |  |  | Humanitarian data exchange  <https://data.humdata.org/> | | Roof and wall types for housing | Type of roof/wall material for houses. Physical vulnerability | Based on 2015 national census | Aggregated per municipality | Philippines National Census <http://psada.psa.gov.ph/index.php/home> | | Damage and Loss data for historical typhoon events | Total number of completely/partially damaged houses | for each event | Aggregated per municipality | National DRR and Management Council (NDRRCM) | | Settlements at risk of Storm surge flood | Number of settlements at high risk of flood due to storm surge |  | Aggregated per municipality | Nationwide Operational Assessment of Hazards (NOAH)<http://noah.up.edu.ph/#/> | | Settlements at risk of Landslide | Number of settlements at high risk of landslide |  | Aggregated per municipality | Nationwide Operational Assessment of Hazards (NOAH)<http://noah.up.edu.ph/#/> | | Topography | Slope, Aspect, Elevation, length of coast line, ruggedness etc.… |  |  | GIS analysis using SRTM elevation data | | Social vulnerability | Poverty, demography | Based on 2015 national census | Aggregated per municipality | Philippines National Census <http://psada.psa.gov.ph/index.php/home> | | Settlement layer | High Resolution Settlement Layer (HRSL) |  | 10 x 10 | <https://www.ciesin.columbia.edu/data/hrsl/> |     **Hazard:** Historical rainfall from NOAA, and historical typhoon tracks from IBTrACS .   * 1. Is it representative of the population being sampled?   The training set fully covers the target population in terms of geographic scope, as the training data is for the entirety of the Philippines while OCHA is interested in regions 5 and 8.   * 1. How accurate or reliable is the training data?   Past typhoons, rainfall, and historical damage data are added after each event and the model retrained.  Many datasets are static such as the total number of houses, housing materials, and poverty, and may not be representative of the current situation.   * 1. How is missing data treated? (e.g., exclusion, single imputation, multiple imputation)   The list of historical typhoons is thought to be complete down to category 2 storms since 2006. Zero and null wind fields are added to the training set to represent the no-damage case.   1. What are the model assumptions and approximations?   The input variables depend on the following assumptions:   * reported damage/loss data is complete for all 40 typhoon events * all municipalities affected by any of these typhoon events were accounted for in each damage reports compiled by NDRRCM * aggregation of input variables per municipality is sufficient to obtain representative values * Static datasets are representative of full timespan (e.g. number of houses, building materials, poverty level etc)   The following approximations are made:   * Impact is aggregated to the municipality level, as lower admin levels on damage/output are unavailable * Only totally damaged houses are considered, and it is assumed that these are correlated with partially damaged houses * The model prediction of completely damaged houses is used as a proxy for severity if impact and number of people affected  1. Methodology - provide a description of the different analysis steps and how the input datasets are used to train the model.   **FIgure 1:** Methodology  3. Model Evaluation   1. Are there other similar existing models and how does this model compare?   A damage model developed by the University of the Philippines institute of civil engineering (UP-ICE) is used as a baseline model to evaluate the performance of 510 typhoon impact prediction models.  **Figure 2:** UPD\_ICE damage curve, For each building type showing the relationship of fraction of damage with wind speed  The UP-ICE model is a physical vulnerability model (Figure 2), which is a curve depicting a relationship between wind speed and building damage, and it is used to estimate damage per building type due to wind.    **Figure 3.** Comparison of performance of best model against Baseline, baseline: UP-ICE baseline model; ranger: best model based on random forest algorithm   1. Have there been any past peer reviews, such as published papers or other review exercises?   The model is used to trigger the typhoon Early Action Protocol(EAP) of the Philippines Red Cross. The EAP approval was an iterative process which included evaluation of the model and its documentation, and was performed by the EAP validation committee of the International Red Cross Federation.   1. Is there any reference or benchmark used to evaluate the performance?   The baseline model uses wind speed plus housing materials applied to damage curves to estimate the percentage of totally damaged houses. It is found to underestimate damage compared to this model.  **Figure 4a**: Baseline model using damaged curves, which is found to underestimate the number of totally damaged houses.    **Figure 4b**: Performance of the current model   1. What are the metrics used for model evaluation? Why have these metrics been selected? Describe current model performance.     **Figure 5:** Model performance   1. How does performance depend on forecast lead time?   The effect of lead time on model performance is assessed indirectly by calculating track error for different lead times. The result is summarized in the table below.   | ***Model*** | *Average Track Error (kms) for different lead times (h)* | | | | | | | | | | | | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | *6* | *12* | *18* | *24* | *30* | *36* | *42* | *48* | *54* | *60* | *66* | *72* | | ***ECMWF*** | *50* | *66* | *73* | *98* | *101* | *134* | *135* | *174* | *172* | *214* | *209* | *255* |  1. What is the risk tolerance of the model? What happens if the model produces false positives, false negatives?   No formal study on the risk of false alarms has yet been performed  4. Operational Readiness   1. Is the model ready to be used to inform humanitarian response?    1. Is the model kept up-to-date with the latest datasets?    2. Who is responsible for model updating and/or recalibration?    3. Is the model ready to be deployed? If not, what are the additional steps needed (e.g. further research, validation, updates)?   The model is already applied to trigger humanitarian early action of the Philippines Red Cross (PRC) within the framework of the Typhoon Early Protocols (EAP) approved in November 2019. The model was applied for typhoon Tisoy (international name Kammuri), end of November 2019, starting 5 days in advance until impact, and early actions were triggered and successfully implemented 2 days before landfall, with financial support from the International Federation of Red Cross and Red Crescent (IFRC) Disaster Relief Emergency Fund (DREF).  Currently 510 works on maintenance, monitoring and improvement of the model. The model is run manually if there is an active typhoon in the Philippines area of responsibility (PAR). The preprocessing scripts rely on information from Global disaster alert and coordination system (GDACS) to identify active typhoons.   1. Has the model been developed in collaboration with operational partners? Has the model previously been used in humanitarian situations?   The FBF project team in the Philippines Red Cross was involved in model development and data collection.  The trigger methodology was tested first with 510 model for typhoon Ondoy (Mangkhut) in September 2018, then two times in 2019, for typhoon Tisoy (Kammuri) - leading to the activation of early actions - then for Ursula (Phanfone), which didn’t trigger any humanitarian early actions as the threshold was reached only less than 12 hours before landfall. However, note that this version of the model used a percentage damage threshold trigger, rather than total number of houses damaged.   1. What happens if the model is inaccurate, produces false positives or false negatives? Describe expected impact according to the in-scope use cases highlighted in Section 1A above.   OCHA is aware of possible prediction errors from trigger models and these are acceptable upto a certain degree. Post performance evaluation will be conducted after any trigger, and levels can be adjusted if needed. In addition, data from the new event will be added to the training set and will be used to re-train the 510 typhoon model.   1. List additional considerations for the use of the model in humanitarian response.   The current model training dataset contains several country-specific variables. However, a variable importance assessment found that the most important variables are related to the typhoon characteristics, which indicates that there is a strong potential to create a more generic model applicable in other locations. |
|
|
|
|
|
|
|
|
|
|