

TOWARDS PREDICTING RICE LOSS DUE TO TYPHOONS IN THE PHILIPPINES

S. Boeke^{1,4}, M.J.C. van den Homberg¹, A. Teklesadik,¹ J. L. D. Fabila², D. Riquet³, M. Alimardani⁴

¹ 510, an initiative of the Netherlands Red Cross

² Philippine Red Cross

³ German Red Cross

⁴ Tilburg University

Commission IV

KEY WORDS: Typhoons, Philippines, Machine Learning, impact-based forecasting, rice loss

ABSTRACT:

Reliable predictions of the impact of natural hazards turning into a disaster is important for better targeting humanitarian response as well as for triggering early action. Open data and machine learning can be used to predict loss and damage to the houses and livelihoods of affected people. This research focuses on agricultural loss, more specifically rice loss in the Philippines due to typhoons. Regression and binary classification algorithms are trained using feature selection methods to find the most important explanatory features. Both geographical data from every province, and typhoon specific features of 11 historical typhoons are used as input. The percentage of lost rice area is considered as the output, with an average value of 7.1%. As for the regression task, the support vector regressor performed best with a Mean Absolute Error of 6.83 percentage points. For the classification model, thresholds of 20%, 30% and 40% are tested in order to find the best performing model. These thresholds represent different levels of lost rice fields for triggering anticipatory action towards farmers. The binary classifiers are trained to increase its ability to rightly predict the positive samples. In all three cases, the support vector classifier performed the best with a recall score of 88%, 75% and 81.82%, respectively. However, the precision score for each of these models was low: 17.05%, 14.46% and 10.84%, respectively. For both the support vector regressor and classifier, of all 14 available input features, only wind speed was selected as explanatory feature. Yet, for the other algorithms that were trained in this study, other sets of features were selected depending also on the hyperparameter settings. This variation in selected feature sets as well as the imprecise predictions were consequences of the small dataset that was used for this study. It is therefore important that data for more typhoons as well as data on other explanatory variables are gathered in order to make more robust and accurate predictions. Also, if loss data becomes available on municipality-level, rather than province-level, the models will become more accurate and valuable for operationalization.

1. INTRODUCTION

Every year, natural hazards like typhoons, earthquakes and wildfires affect approximately 160 million people worldwide (WHO, 2019). Due to climate change it is expected that scale, frequency and impact of these hazards will increase over the coming years which will result in an increasing demand of humanitarian aid (Ashdown, 2011). However, not all of these people can be helped due to limited funding and resources for disaster response. Therefore, the areas and individuals with the highest priority have to be determined. During this process, humanitarian organizations quickly need insight on the damage in the areas that were hit by the disaster. Nevertheless, this can often be very time consuming and sometimes also subjective because of a lack of data and tools to interpret them (van den Homberg et al., 2018; van den Homberg, 2017). Most of the disaster funding in the last 20 years was spent on emergency response, reconstruction, and rehabilitation (Kellett, Caravani, 2013). According to a study of disaster-related financing by the Global Facility for Disaster Reduction and Recovery (GFDRR) and the Overseas Development Institute (ODI), only 13% is invested in reducing the risk of disaster before it happens. That such a small amount is spent on the pre-disaster phase is not very cost-effective, according to the review study of Mechler (Mechler, 2005). In the majority of evaluations of preventative action, the avoided disaster losses could at least double the investment in risk reduction. Nobre et al. also found positive results regarding preventative actions. In their study the potential cost-effectiveness of cash transfer responses were evaluated, comparing the relative costs of ex-ante cash transfers during the maize growing season to ex-post cash transfers after harvesting in Kenya (Nobre et al., 2019). Overall, their findings suggest that early response can yield significant cost savings, and can potentially increase the effectiveness of existing cash transfer systems. Although this study is focused on crop damages due to extreme drought, it could be inferred that ex-ante cash transfers can also be helpful in cases of other weather and climate hazards. Given the above positive cost-benefit ratios, humanitarian funding is more and more becoming available in the pre-disaster phase to trigger preventative actions in the time

window between a weather forecast and a potential disaster (Wilkinson et al., 2018). Humanitarian organizations such as FAO and the Red Cross Red Crescent Movement (DRK, 2019), are transforming their way of operating and are developing trigger models for early action. This transformation requires a data-driven approach that connects data on different risk predictors with actual impact. Machine learning can provide new methods of looking into these connections and provide more accurate and useful answers (GFDRR, 2018). In the private sector such as retail and finance, several machine learning techniques have been developed to deal with the greater amounts of (big) data becoming available. More and more of these commercially developed machine learning techniques are applied for social good, dealing with bigger issues confronting humans, like preparing for and recovering from disasters (GFDRR, 2018). About 20 years ago, machine learning techniques were introduced in the field of disaster management and have become one of the most effective methods for removing unrelated data and speeding up the analysis in disaster situations, which helps in fast prediction analysis and finding optimal response approaches (Yang, Li, 2018). Caragea et al. (2011) classified tweets and text messages to address the most urgent needs and to better understand the Haiti earthquake emergency situation.

510, the humanitarian data initiative of the Netherlands Red Cross, developed a priority index that predicts the extent of damage to houses per municipality within 12 hours after a typhoon makes landfall (510Global, 2019). The results can give organizations like the Philippine Red Cross and UN OCHA an overview of the geographic distribution of damage when there are no other sources of reports available yet. Models to predict what the impact will be of an impending disaster are basically the same as models that do so before the disaster hits. Only in the latter case, one has to use hazard forecasts instead of actual hazard data adding more uncertainties (that increase with lead time) to the prediction of the impact. More and more studies use machine learning techniques to predict different types of damage due to natural disasters. Rajasekaran et al. (2008) used support vector regression to predict

storm surges, while Hu and Ho (2014) used the same method to predict the impact of typhoons on transportation networks. Wang et al. (2015) applied a random forest classifier to flood hazard risk assessment and Pradhan and Jebur (2017) used both K-nearest neighbour and logistic regression algorithms to predict regions susceptible to natural hazards. Choi et al. (2018) experimented with algorithms like decision trees and random forests to predict heavy rain damage. Rajasekaran et al. (2008) used a support vector regressor to predict storm surges in order to avoid property loss and reduce risk by taking selective preventative action. As the examples show above, most studies have focused on impact on infrastructure. However, one of the other sectors that is often affected by natural hazards is agriculture. In developing countries, this sector absorbs 23 percent of the total damage and losses. These damages have large negative impacts like disrupting production cycles, trade flows and livelihood means (FAO, 2017). It is therefore important to try to diminish the consequences by improving the preparation for these hazards. This study has focused on the prediction of rice loss due to typhoons in the Philippines. These storms are also known as cyclones or hurricanes, depending on where they occur. Due to its geographical location, the Philippines is very vulnerable to typhoons. Annually, an average of 22 typhoons enter the archipelagic country, and with approximately seven typhoons that cause significant damage it ranks second after China (Division, 2015). Farmers that harvest rice, which is the staple food in the Philippines, often suffer from these hazards. The average damages to rice farming between 2007 and 2010 due to typhoons in the Philippines amounted to 7,996.59 million Philippine Pesos per year (135 million Euro) (Israel, 2012). In order to diminish these damages, the Department of Agriculture (DA) regional office can lend two combine harvesters/threshers to farmers in order to save their crops from impending typhoons. According to the DA, farmers found this to be very advantageous in the past. As for the modelling, until now not more than a handful of studies have focused on the prediction of losses due to typhoons. Both studies of Blanc and Strobl (2016) and Masutomi et al. (2012) have used fragility curves for estimating the extent of crop areas damaged by typhoons. Such a function relates external forces and the probability of damage. Chiang et al. (2012) evaluated the impact of typhoons on agriculture in Taiwan and predicted the agricultural losses by a neural network. One limitation mentioned in the latter study is that the constructed model can only produce the losses at national instead of on a more local level. The same limitations were also addressed in the study of Masutomi et al. (2012). Koide et al. (2012) developed mostly regression type of models to predict rice yield at the end of a growing season with input variables on the cultivation process, rainfall and tropical cyclone activity (as captured in the Accumulated Cyclone Energy). The objective of our study is to determine if and how rice loss due to typhoons can be predicted using open data and machine learning. To this end, we experimented with several machine learning techniques and corresponding feature selections in a case study for the Philippines. In contrast with some of the previous studies regarding crop damage prediction, our models were trained to eventually make predictions a couple of days before the typhoon will make landfall, rather than shortly after the typhoon occurred.

2. MATERIALS AND METHODS

In this section, the data as well as the methods used for the different prediction algorithms will be discussed.

2.1 Case study and data

The models were trained on data of 12 typhoons that hit the Philippines in the past eight years. These typhoons were chosen by 510 because these were the most important typhoons in the past eight years, and most of the in- and output data for these typhoons were used for earlier prediction models of 510 (HDX,

2019 and Wagenaar et al., 2019). The names of these typhoons are: Goni, Kalmaegi, Koppu, Sarika, Haima, Melor, Rammasun, Utor, Nock-Ten, Hagupit, Haiyan and Bopha. Typhoons Sarika and Haima occurred in October 2016 a couple of days one after another, whereby Haima was a much bigger event than Sarika. The Department of Agriculture clustered the rice damage of these two typhoons; we therefore decided to label these two events as Haima, reducing hereby the amount of labelled events to 11. Data can be found on 510's community risk assessment dashboard (CRA, 2019). In the next sections, first the output variable will be discussed, followed by a section regarding the input variables. Finally, in the last subsection, some visualizations and insights on the data will be given.

2.1.1 Output variable

The Philippines Department of Agriculture provided data on the rice damage at province level after filing a Freedom of Information (FOI, 2019) request. The output of the regression models is the percentage of the rice area that has been totally damaged. To obtain the damaged proportion, the rice area and the damaged rice area had to be known. First, the rice area for every province in the Philippines was obtained via Philippines Rice Information System (PRISM). PRISM collects data and generates rice production information using mobile technology, remote sensing, and Geographical Information System (GIS) (PRISM, 2019). PRISM has rice area data on province-level of 2018 and 2017 for both the first and second semester. Semester 2 refers to the first cropping, usually starting around June. Semester 1 refers to the second cropping, which starts around November. The majority of the provinces has more rice area in Semester 2. This could be because farmers usually plant more rice at the end of the dry season (June-July) to benefit from the rain, especially in areas without irrigation systems. For this study, the rice area data of 2018 were used. For every observation, depending on when the typhoon occurred, the rice area for that province corresponding to either the first or second semester was used. Furthermore, a distinction was made in the damage data between totally damaged and with chance of recovery. For this study, only the totally damaged (lost) area was taken into account. After calculating the loss percentage for each observation, three of them exceeded 100%. It could be the case that in that time, there was more rice planted than in 2018. Therefore, these values were replaced by 100%. In total, 224 observations were used to train the models. Next to the three regression models, three binary classification models were trained, each of them having a different threshold. These thresholds are the percentage of the lost rice fields and represent the critical value for triggering early action. As there is not yet a gold standard for preventative action, a threshold of 20%, 30% and 40% was used.

2.1.2 Input variables

The 14 input variables can be sub-divided in two categories: geographical and typhoon features. In Table 1, the features for every category are listed. Data on these input variables is obtained through desk research (scouting global and national data repositories) as well as through visiting organizations in-country (van Lint, 2016). We added for this research data on rice area, total length of streams and rivers and the drainage density. Quantum Geographic Information System (QGIS) was used to analyze and process the geospatial source data. The land map of the Philippines was covered with so called raster layers: matrices of cells that represent features on the earth's surface. Each layer represents a different feature, and every cell in this layer contains the value for this feature (QGIS, 2019). Some of these features were initially on municipality-level, and therefore needed a transformation to province-level. The average elevation, slope and ruggedness, were obtained by taking the area weighted average for all the municipalities in a certain province. Figure 1 shows the variation in the width and height of each province, with each province having an average area of 4000 km².

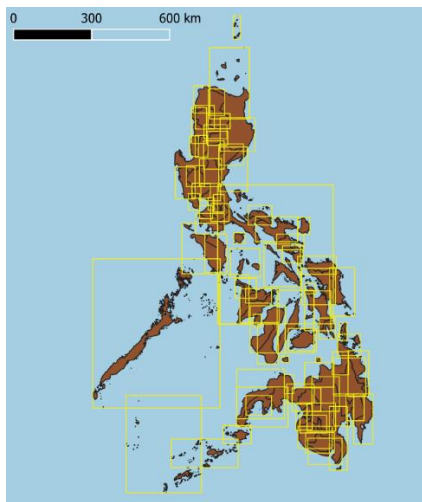


Figure 1. The width and height of the 81 provinces in the Philippines.

The ruggedness is defined as the mean difference between a central cell's elevation value and the elevation values of its surrounding cells. The slope is defined as the angle of inclination to the horizontal. The X and Y coordinates are the transformed latitude and longitude of the centroid of the province. The drainage density is the ratio between the total length of streams and rivers and the total area of the province. The total rainfall is defined as the cumulative rainfall during the typhoon event. Furthermore, the distance to typhoon is the perpendicular distance from the center of a province to the track of that typhoon. The coastal length is zero for a province that is landlocked and gets larger for provinces that are an island; the ratio of the coastal length divided by the perimeter of the province is the coast-perimeter ratio. For three of the 11 typhoons, the rainfall data were incomplete, so a second source was used to fill the missing rainfall data. This source was used for a previous project of 510 and contains data on municipality-level and thus had to be aggregated. This was done by taking the area weighted average. After using this second source, the rainfall data for 17 observations were still not filled. Therefore, the missing entries were replaced by the mean of all the available rainfall data. Worth noting is that the average wind speed was only captured when it was equal or larger than 40 miles per hour (mph). If this was not the case, an average wind speed of 0 mph was noted. As typhoons often have a high wind speed, this probably did not affect provinces close to the typhoon track. Therefore, it was expected that this would not negatively impact the prediction models in a significant way.

2.1.3 Data exploration

As mentioned before, the data consisted of 224 observations. These observations concern 57 different provinces. In Figure 2, the average loss per typhoon is visualized. The total average is 7.1% and was computed by taking the average over all observations. Although Haiyan was one of the strongest typhoons the Philippines ever encountered, one can see that typhoon Nock-Ten caused the biggest rice loss (percentage wise). Further detailed study will be required to understand the reasons for this, although it might be related to the timing of the typhoon in the rice growing season. Nock-Ten hit the rice crops in Bicol province at the end of December when they were in advanced growth stages and got permanently lost (USDA, 2017), whereas typhoon Haiyan made landfall almost two months earlier in the growing season. Hence, the rice loss due to earlier typhoons, like typhoon Haiyan (2013), seem smaller. In Figure 3, one can see that the number of observations vary a lot between typhoons. This was taken into account when evaluating the

models, which will be explained in detail in the next subsection. Furthermore, in Figure 4, the Pearson correlations of the features with rice loss are presented. As expected, the wind speed has the largest correlation ($r=0.41$, $=1.78e-5$). The distance from the typhoon has the largest negative correlation ($r=-0.29$, $=0.002$).

Geographic	Typhoon
Total area	Average wind speed
Total rice area	Total rainfall
Average elevation	Distance to typhoon
Average slope	
Average ruggedness	
Coast length	
Coast-perimeter ratio	
X and Y coordinates	
Total length of streams and Rivers	
Drainage density	

Table 1. Input variables

2.2 Machine learning models

The algorithms in this study were trained and tested on a MacBook Air (2017) with an Intel Core i5 1.8-GHz processor, 8 GB 1600 MHz DDR3 RAM. The programming language used is Python 3.6.8, using the software libraries Scikit-learn version 0.20.3 and Pandas version 0.24.1. The code developed is available on the <https://github.com/rodekruijs>.

2.2.1 Overall approach

In this section, the regression and classification models that were used for prediction are discussed in detail. In Table 2 the algorithms per task are listed as well as the features used. We will now discuss the evaluation and feature selection methods of the models.

Evaluation To evaluate every model's performance, an 11-fold cross validation was carried out. For every fold, all the observations that concern one of the 11 typhoons were left out as a test set. This method is the most realistic, as in a real situation the damage of a typhoon is also being predicted on the basis of previous typhoons. Because of the inexplicable typhoon specific variation in the output, the performance of the model would be biased if both the training and test set contained observations of a certain typhoon. Finally, the average of the 11 performance metrics was taken. However, as there was a big variation in the number of observations per typhoon (see Figure 3), the weighted average was taken. The evaluation metrics that were used for the regression and classification models are discussed later on.

Feature Selection For every separate algorithm, feature selection was carried out. This means that only the most important features are taken into account, as sometimes too many features can lead to poor generalization of the model (Guyon and Elisseeff, 2003). It is important that this is done together with the cross validation procedure, in order to prevent bias of the evaluation metric. If first the best features are selected based on all the data, and only thereafter cross-validation is used to estimate the prediction error, then the test set is not completely independent. This means that the test data in every fold of the cross-validation procedure would also be used to choose the best features. Feature selection is actually a kind of training and training can never be done on the test set as well. Therefore, test samples must be left out before selection steps are carried out (Friedman et al., 2001). As there are different approaches to feature selection, this study has experimented with different modules in order to find the best performing feature set for each algorithm. Every model has tried each of the following two approaches; Recursive Feature Elimination and SelectKBest.

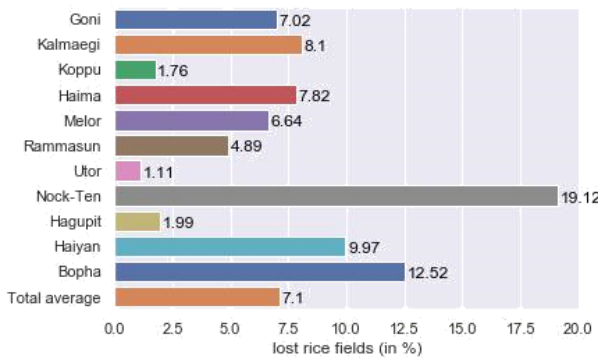


Figure 2 Average rice loss

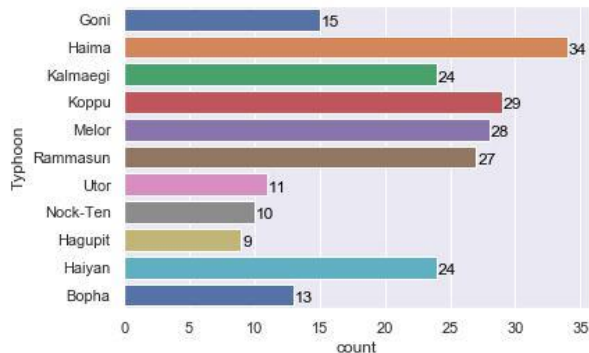


Figure 3. Observations per typhoon

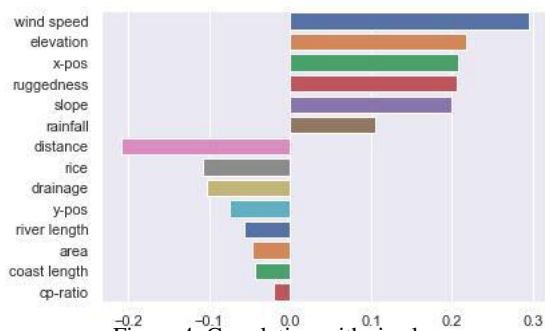


Figure 4. Correlation with rice loss

First of all, Scikit-Learn's module Recursive Feature Elimination (RFE) was used. In this case, the estimator starts with all the initial features. Then recursively, the least important feature is being discarded until the desired number of features is reached. The importance of a feature is measured by the estimator's rise in error or drop in accuracy, when this feature is being discarded. In contrast to the next method, RFE does not support the use of a support vector machine/regressor and k-nearest neighbor algorithm. This is because there is no notion of feature importance given in these algorithms. Therefore, this method was only used when training a random forest model and a linear regressor. This is because there is no notion of feature importance given in these algorithms. Therefore, this method was only used when training a random forest model and a linear regressor. Secondly, the selection method SelectKBest from Scikit-Learn's univariate feature selection approach was used. Here, statistical tests are used to select k features that have the strongest relationship with the output variable. For the regression models, the scoring functions that were used are f regression and mutual info regression. As for the classification models, chi2, f_classif and mutual_info_classif were used as scoring functions. Both the f_regression and f_classif tests rank features according to their correlation with the output variable. The mutual info tests are non-parametric, meaning that there are no distribution assumptions, which use k-nearest neighbors to measure the degree of relatedness between the features and the output variable (Ross, 2014). Finally, the chi2 test computes the chi-

squared statistic between each non-negative feature and class. The chi-square test measures dependence between stochastic variables, so using this function leaves out the features that are the most likely to be independent of class and therefore irrelevant for classification (Pedregosa et al., 2011). The number of features, the selection method, and the hyperparameters are interdependent for each model. Therefore, the number of features and the feature selection method were included in the grid search for every separate algorithm. In practice, this was carried out through nested for-loops, such that every possible combination of hyperparameters, number of features and selection methods was evaluated. For some algorithms, the features that were selected were not the same for every fold. In this case, the features that had the maximum number of votes were chosen and the cross validation procedure, including a new hyperparameter search, was repeated with these features.

2.2.2 Regression

For the regression algorithms, the Mean Absolute Error (MAE) was used to evaluate the performance. This easily interpretable metric was preferred over Root Mean Squared Error (RMSE) and Mean Squared Error (MSE) as larger errors did not necessarily need to be more penalized in this study. As mentioned before, after applying 11-fold cross-validation, the MAE was determined by taking the test size-weighted average of all the MAE's. After tuning the algorithm's hyperparameters, a baseline was defined by running that same model but only with wind speed as explanatory variable, as this feature has the biggest correlation with rice loss. Setting this baseline was only to put the model's performance into perspective. If the baseline would perform very well, this was also a positive outcome. Now, the different regression models and their hyperparameter settings will be discussed.

Multiple linear regression First of all, a multiple linear regression model was implemented. One of the reasons is that this model has the ability to make a prediction that is higher than the values in the training data. This is called extrapolating for when an extreme typhoon occurs that was not seen earlier in the training data. For this straight forward algorithm, no hyperparameters were tuned. As for the feature selection, the f regression test selecting only wind speed as explanatory variable led to the lowest MAE.

Random forest Second, a random forest regressor was used to predict the lost rice area. In contrast with linear regression, random forest models are able to detect non-linear relations. However, if applied to extrapolating domains, it could lead to poor predictions (Hengl et al., 2018). In a random forest, multiple decision trees are generated, where after the random forest predictor is formed by taking the average over the trees. Each of these trees uses a random subset of features and a new training set. This training set is drawn, with replacement, from the original training set, also known as bagging (Breiman, 2001). There are several hyperparameters that can be tuned for this algorithm. The optimal values for the number of estimators (n_est), the maximum depth (max_depth), the minimum sample split (min_split) and the minimum sample leaf (min_leaf) were found through grid search. The other hyperparameters were kept at their default values. The number of estimators represents the number of trees used in the forest. The maximum depth represents the maximum depth of every tree in the forest. The deeper the tree, the more information it can extract from the data. The minimum sample split stands for the minimum number of samples required to split an internal node. The minimum sample leaf parameter specifies the minimum number of samples in a leaf node (last node of the tree). Setting it lower leads to trees with a larger depth which means that more splits are performed until the leaf nodes (Probst et al., 2018). The search ranges were initialized to be [4; 8], f2; 4; 6; 8; 10g, [1; 5] [f0:1g and [1; 5] [f0:1g for n_est, max_depth, min_split and min leaf, respectively. For min_split and min_leaf, 0.1 represents the fraction of samples. A total of five features were selected with the

use of RFE. These were not identical for every fold. The most common features were wind speed, rice area, drainage density, average slope and average elevation. After repeating the cross validation procedure with these features, the following hyperparameter values obtained the lowest MAE: $n_est=4$, $max_depth=4$, $min_split=2$ and $min_leaf=4$.

Support vector regression Finally, a support vector regressor (SVR) was used. This choice was based on several studies that have predicted the impact of typhoons using this algorithm (Hu and Ho, 2014). SVR supports both linear and non-linear regression tasks. The goal is to find a function f that has at most ϵ deviation of the actual targets y_i (Smola and Scholkopf, 2004). However, such a function does not always exist. Therefore, the penalty parameter C is introduced, which determines the trade-off between the algorithm's complexity and the amount up to which deviations larger than are tolerated. The kernel coefficient gamma contrast to the next method, RFE does not support the use of a defines how far the influence of a single training example reaches. When gamma is too small, the model is too constrained and cannot capture the complexity of the data. For gamma, C and the ϵ hyperparameters tried are $\{ 'auto', 2e-4, 1e-4, 0.001, 0.1, 1 \}$, $\{ 0.001, 0.1, 1, 2 \}$ and $\{ 0, 0.001, 0.1, 0.4, 0.8, 0.9 \}$ respectively. When gamma='auto', the inverse of the number of features is taken. For the SVR, the $f_regression$ test selected just one feature: wind speed. The values of the hyperparameters that led to the best generalization were: $\epsilon = 0.4$, $C=1$ and $gamma=0.001$.

2.2.3 Classification

In this subsection, the algorithms that were used for binary classification are discussed. Every algorithm was trained three times, each of them with a different threshold: 20%, 30% or 40%. The numbers are the rice loss percentages and represent the critical level for triggering early action. For these classification models, the accuracy, recall, precision and f1-score were evaluated. As mentioned before, after the cross-validation procedure, the weighted average of the evaluation metrics was taken. The accuracy was test size-weighted. The recall was weighted on the basis of the number of positive samples in y_{true} . As for the precision, this metric was weighted on the basis of the number of positive samples in y_{pred} . For some folds, either the precision or recall was undefined (weight 0) such that the f1-score could not be computed. The developers of f-score have experimented with different methods when dealing with this problem. They recommended numbering the number of true positives (TP), false positive (FP) and false negatives (FN) over the folds and then use equation (1), as this method is almost perfectly unbiased. Furthermore, as in all the three cases the classes were imbalanced (see Figure 5), a re-sampling method was applied. In detail, a random oversampler aimed to balance class distribution through the random replication of minority class examples (Batista et al., 2004). It is important to apply such a re-sampling method only after the test set is split from the training set. If first re-sampling is applied and then the data are split, both the training and test set could contain the exact same observations, resulting in a bias of the evaluation metric. Now, the different classification models and their hyperparameter settings will be discussed.

$$F_1 \text{ score} = 2 \frac{Pr \cdot Re}{Pr + Re} = 2 \frac{\left(\frac{TP}{TP+FP}\right)\left(\frac{TP}{TP+FN}\right)}{\left(\frac{TP}{TP+FP}\right) + \left(\frac{TP}{TP+FN}\right)} = 2 \cdot \frac{TP}{TP + FN + FP} \quad (1)$$

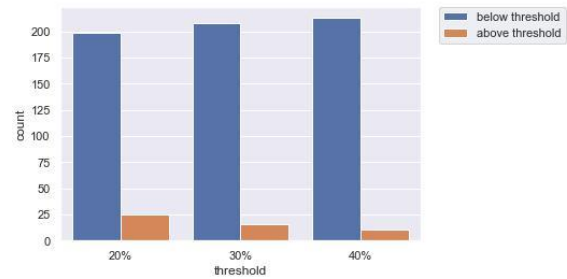


Figure 5. Class distribution for different thresholds

Random forest Next to regressors, random forest models can also serve as classifiers. Instead of taking the average of the outcomes of every decision tree, the classifier assigns the observation to the majority vote class. The hyperparameters and their search ranges used for the random forest regressor, were also used for the classification task. For the binary classification where the threshold was set at 20%, the RFE method selected 3 features: wind speed, drainage density and rice area. The following hyperparameters performed best: $n_est=6$, $max_depth=2$, $min_split=0.1$ and $min_leaf=0.1$. For the threshold of 30%, the model performed best when a total of 9 features were included. The features that were not included are: total area, coast-perimeter ratio and the X and Y coordinate. The following hyperparameter values were obtained: $n_est=4$, $max_depth=2$, $min_split=0.1$ and $min_leaf=0.1$. For the model with a threshold of 40%, the following three features were selected with the RFE method: wind speed, slope and drainage density. Using these 3 features, the model performed best with the following hyperparameter values: $n_est=4$, $max_depth=4$, $min_split=0.1$ and $min_leaf=4$.

Support vector machine Furthermore, a support vector machine (SVM) classifier was trained. The same hyperparameters as for the support vector regressor were tuned, except for ϵ . This hyperparameter is nonexistent for classification, as the aim is not to create a function that predicts the target value but to find a hyperplane that maximizes the margin between the two classes. In addition, the tolerance (tol) of the stopping criterion was also taken into account during the grid search, as this hyperparameter substantially increased the model's performance. For the tolerance, the set $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ g was used for grid search. For the other hyperparameters, the hyperparameter sets were equal to the ones of the support vector regressor. In all three cases, only wind speed was selected as feature with the $f_classif$ test. For the model with a threshold of 20%, the hyperparameter values that performed best were $C=0.001$, $gamma='auto'$ and $tol=0.1$. When the threshold was set at 30%, the following hyperparameter values were selected: $gamma=1e-4$, $C=0.001$, $tol=0.1$. Finally, for the model with a threshold of 40% the following hyperparameter values were found through grid search: $C=0.001$, $gamma=2e-4$ and $tol=0.1$.

K-nearest neighbour Finally, a k-nearest neighbour (k-NN) classifier was trained. This algorithm assigns every observation from the test set to the class of majority vote of the nearest k neighbors (Cover, Hart, 1967). The value of k is a hyperparameter for which the search range was set at $[1; 15]$. For the classification with a threshold of 20%, two features were selected with the use of the $f_classif$ test: wind speed and distance to typhoon. The optimal value for k was 15. For the model with a threshold of 30%, a total of 4 features were selected with the use of the mutual info $f_classif$ test: distance to typhoon, total rainfall, slope and rice area. Setting the value of k at 11 made the model perform best. Finally, the $f_classif$ test selected one feature for the model with a threshold at 40%. Not all selected features were the same, but wind speed was the most common feature. After repeating the grid search, an optimal value of 11 was found for k.

Machine learning model	Algorithm (threshold)	Features
Regression	Linear regression	Windspeed
	Random forest regressor	Wind speed, rice area, drainage density,
	Support vector regressor	Windspeed
Classification	Random forest classifier (20%)	Wind speed, drainage density and rice area
	Random forest classifier (30%)	Average wind speed, Rice area, Average elevation, Average slope, Average ruggedness, Coast length, Total length of streams and rivers, Drainage density, Total rainfall, Distance to typhoon
	Random forest classifier (40%)	Wind speed, slope and drainage density
	Support vector classifier (20, 30 and 40%)	Wind speed
	K-nearest neighbour (20%)	Wind speed and distance to typhoon
	K-nearest neighbour (30%)	Distance to typhoon, total rainfall, slope and rice area.
	K-nearest neighbour (40%)	Wind speed

Table 2. Overview of machine learning models: algorithms and features used

3. RESULTS

In this section, the results of the regression and classification models are presented. The outcomes of the regression algorithms are given in percentage point. For every task, the algorithm in bold is the best performing algorithm. One can see that the support vector regressor performed best at the regression task with an MAE of 6.83 percentage points, representing the average difference between the predicted and the actual value. However, with an MAE of 7.36 percentage points the random forest regressor did not perform significantly worse. The linear regression model performed relatively worst, with an MAE of 8.90 percentage points. As for the binary classifiers, the support vector machine obtained the highest recall score for all three thresholds. Recall scores of 88%, 75% and 81.82% were obtained when the threshold was set at 20%, 30% and 40%, respectively. One can clearly see the trade-off between the recall and precision of every classification model. When more positive labels are predicted, chances are high that most of the positive samples are rightly predicted, resulting in a high recall score. However, there is also a big chance that the predictor falsely labels the negative samples as positive, which causes a low precision score. Acting on these outcomes, would mean provinces receive aid although they are not in need of aid.

4. DISCUSSION

Our study combined open data with regression and classification machine learning algorithms to determine how rice loss due to typhoons can be predicted for a case study in the Philippines. The regression models were trained to improve the MAE and the classification models were trained to improve the recall score. It is considered more important to rightly classify the provinces that are in need of aid, than to rightly classify the provinces that are not. Missing out on farmers that are in need is usually considered as a more serious error than acting in vain. But only a thorough valuation approach can provide an answer to what is the right balance (Lopez et al., 2018). The support vector regressor performed best with a MAE of 6.83 percentage points. As for the binary classification, the models were evaluated on their ability to rightly classify the positive samples. For all three thresholds, the support vector classifier obtained the highest recall score. For every threshold, the precision score was high. However, for all of these models the precision score was low. For both regression and classification tasks, the support vector machine performed best. For each of these support vector machine models, only wind speed was used as explanatory feature. It is remarkable that for most of the other algorithms, different sets of features came out best. Also, for the random forest algorithms, the hyperparameters used for the algorithms had great impact on features that were selected. This suggests that with the data that was available for

this study, there was no convincing outcome regarding the most important features when predicting rice loss. There are a number of limitations of this study that could be improved in future research. The first and most important limitation is the amount of data that was available. It is important that more data are gathered such that more robust and accurate predictions can be made. With only 224 observations, this study has given an insight on the possibilities of rice loss prediction but the estimations are imprecise. Not only were there not much training data to train the models on, as an 11-fold cross-validation was used, each predictor was only tested on a test set consisting of only 20 instances on average, which does not give a thorough reflection of the prediction performance. As already stated above, the most important features were not similar across the different algorithms and for some algorithms the hyperparameter settings had great impact on the features that were selected. It is expected that when there are more data available to train the models on, it will become clearer which features are most important when predicting rice loss. However, it is important to keep in mind that the set of most important features may differ per country. For example, it could be the case that for a certain country with a varied vegetation, the tree cover density is an important feature (shielding rice fields from the wind) whereas for another country it is not. However, a feature like wind speed is likely to be of great importance in all countries. Next to the amount of data, also the quality of the data can be improved. For example, the rainfall data were not complete and the total rice area data from 2018 were used although the typhoons occurred between 2012 and 2016. In reality the rice area, and thus the percentage rice loss, could have been larger or smaller. Also, the fact that the models in this study were on province-level rather than on municipality-level could have played a role in the mediocre performance of some of the models. It is expected that the prediction models on municipality-level would perform better, as the input features would be more specific, instead of the global, aggregated province features.

From the disaster management perspective, it would also be more valuable when predictions would be on municipality-level rather than on province-level, as in the latter case the outcome does not give a thorough insight of where exactly the early action or aid is needed. On average, a province of the Philippines has an area of almost 4000 km², with considerable differences in lengths and widths as shown before in Figure 1. The average north-south length is 115 km. This means that the wind speed (the parameter most correlated with damage) varies a lot when a typhoon traverses a province (generally from east to west rather than north to south). This supports the need for municipal-level information and analysis. It is therefore important that a similar study, but then on municipality-level, will be carried out. Furthermore, the models in this study were trained with the objective to eventually predict rice loss a couple of days before a typhoon would make landfall. This means that the typhoon related features themselves

are also obtained through a prediction model, such as ensemble forecasting models from University College London.

Regression		
Model	MAE (p.p.)	Baseline (p.p.)
Linear Regression	8.90	8.90
Random Forest	7.36	8.65
SVR	6.83	6.83

Classification with threshold of 20% loss				
Model	Acc (%)	Rec (%)	Prec (%)	F1 (%)
RF	81.25	68.00	33.33	57.63
SVM	50.89	88.00	17.05	33.33
k-NN	76.79	84.00	30.43	57.53

Classification with threshold of 30% loss

Model	Acc (%)	Rec (%)	Prec (%)	F1 (%)
RF	72.32	68.75	16.18	26.20
SVM	66.52	75.00	14.46	27.59
k-NN	70.09	56.25	13.04	23.68

Classification with threshold of 40% loss

Model	Acc (%)	Rec (%)	Prec (%)	F1 (%)
RF	83.93	72.73	19.51	36.36
SVM	66.07	81.82	10.84	21.18
k-NN	79.91	72.73	16.00	30.19

Table 3. Results

Forecasts on typhoon features are available for example at 72, 48 and 24 hours' time intervals before the typhoon makes landfall, whereby the uncertainty decreases. This uncertainty negatively influences the final performance of the impact model. Finally, the use of additional features could result in a better performance of the models. For example, in their study, Blanc and Strobl (2016) made a distinction between irrigated and rain-fed rice fields. Their findings suggested that irrigation technology is better able to deal with the potential damage due to typhoons. For example, irrigation systems may be able to counteract the excessive flooding during a storm. Chiang et al. (2012) used a neural network to predict typhoon induced losses on agriculture. They used among others, the minimum atmospheric pressure and the coverage of the typhoon as explanatory features. If these above measures make indeed the rice crop prediction model better, then the model can be discussed with implementing stakeholders. It will be essential to agree on levels beyond which the model will trigger early action, whereby these trigger levels can be dependent on the type of early action to take (such as early harvesting, irrigation or cash transfers). The Fairness, Accountability, and Transparency in Machine Learning community proposes five principles that will help those involved in early warning early action to implement algorithms in an accountable way following five principles (FAT/ML, 2019): Responsibility, Explainability, Accuracy, Auditability and Fairness. It will also be necessary to describe the use of a model within the larger scope of an early action protocol (which includes for example also an assessment of the different early actions possible). Finally, it is important to note that The National Disaster Risk Reduction and Management Council of the Philippines (NDRRMC, 2019) revised mid 2019 the guidelines for declaring a state of calamity. The use of science-based forecasts are allowed to trigger faster utilization of funds at municipality level, whereby it is no longer necessary to have a whole province to be affected before declaring a state of calamity is possible.

5. CONCLUSION

To conclude, this study has tried different machine learning techniques to predict rice loss due to typhoons in the Philippines. This study has been carried out because it is expected that a lot of damages that farmers encounter as a result of these typhoons can be saved when preventative actions are taken. For this task, the data of 11 typhoons were used to train both regression and binary classification models on province-level. For the binary classification, three different thresholds that represent the critical value for triggering early action were used: 20%, 30% and 40% lost rice area. The classifiers were trained to increase its ability to rightly predict the damaged areas. For the regression task, as well as for the three different binary classification tasks, the support vector machine performed best but were still imprecise. For each of these latter models only wind speed was selected as explanatory feature. However, for the other algorithms that were trained in this study, other sets of features were used. Moreover, for some algorithms the hyperparameter settings had great impact on the features that were selected. This is most probably a consequence of the small data set that was used for this study. It is therefore important that more data are gathered in order to make more robust and accurate predictions in the future. Also, if there can be damage data collected on municipality-level, rather than province-level, it is expected that the models would be more accurate and valuable.

ACKNOWLEDGEMENTS

We would like to thank the Philippine Department of Agriculture for providing us with their valuable insights and data in relation to rice crop and damage. The German Red Cross and the Princess Margriet Fund supported this work financially.

REFERENCES

- 510.Global, Faster response with capacity building Philippines part 1, <https://www.510.global/surge-informatio-the-philippines/>, retrieved Sept 2019.
- Ashdown, P., 2011. Humanitarian emergency response review.
- Batista, G. E. A. P. A., Prati, R. C., Monard, M. C., 2004. A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD explorations newsletter, 6.1, 20-29.
- Blanc, E., Strobl, E., 2016. Assessing the impact of typhoons on rice production in the Philippines. Journal of Applied Met-eorology and Climatology, 55.4, 993–1007.
- Breiman, L., 2001. Random forests. Machine learning, 35-1, 5-32.
- Caragea, C., McNeese, N., Jaiswal, A., Traylor, G., Kim, H., Mitra, P., Wu, D., Tapia, A. H., Giles, L., Jansen, B. J., Yen, J., 2011. Classifying text messages for the Haiti earthquake. Proceedings of the 8th international conference on information systems for crisis response and management, 8.5.
- Chiang, Y., Cheng, W., Chang, F., 2012. A hybrid artificial neural network-based agri-economic model for predicting typhoon-induced losses. Natural hazards, 63.2, 769-787.
- Choi, C., Kim, J., Kim, J., Kim, D., Bae, Y., Kim, H., 2018. Development of heavy rain damage prediction model using machine learning based on big data. Advances in Meteorology.
- Coughlan de Perez, E., van den Hurk, B., van Aalst, M. K., Jongman, B., Klose, T., Suarez, P., 2015. Forecast-based financing: an approach for catalyzing humanitarian action based on extreme weather and climate forecasts. Natural Hazards Earth System Sciences, 15(2012), 895+.
- Cover, T. M., Hart, P. E., 1967. Nearest neighbor pattern classification. IEEE transactions on information theory, 13.1, 21-27.

- Division, H. R., 2015. Subject: E25) Which countries have had the most tropical cyclones hits? Frequently Asked Questions. <http://www.aoml.noaa.gov/hrd/tcfaq/E25.html>.
- CRA 2019, https://dashboard.510.global#!/community_risk?country=PHL
- DRK, G. R. C., 2019. What is Forecast-based Financing? <https://www.forecast-based-financing.org/>.
- FAO, F. A. O., 2017. The impact of disasters on agriculture: Addressing the information gap.
- FAT/ML, 2019, <https://www.fatml.org/resources/principles-for-accountable-algorithms>
- FOI, 2019. <https://www.foi.gov.ph/requests>, ID: #DA-54466385434
- Friedman, J., Hastie, R., Tibshirani, T., 2001. The elements of statistical learning: Volume 1. Springer, New York.
- GFDRR, 2018. Machine Learning for Disaster Risk Management.
- Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. *Journal of machine learning research*, 1157–1182.
- HDX 2019, <https://data.humdata.org/dataset/philippines-typhoon-mangkhut-priority-index-forecasted-damage>
- Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B., Graler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6, 1578–1587.
- Hu, T. Y., Ho, W. M., 2014. Prediction of the Impact of Typhoons on Transportation Networks with Support Vector Regression. *Journal of Transportation Engineering*, 10.
- Israel, D., 2012. Typhoons, floods, and droughts: Regional occurrence and value of damages to rice farming in the Philippines. *PIDS Policy Notes*, 1.
- Kellett, J., Caravani, A., 2013. Financing Disaster Risk Reduction: A 20 year story of international aid. ODI, <http://www.odi.org.uk/publications/7452-climate-finance-disaster-risk-reduction>.
- Koide, N., Robertson, A. W., Ines, A. V., Qian, J. H., DeWitt, D. G., & Lucero, A. (2013). Prediction of rice production in the Philippines using seasonal climate forecasts. *Journal of Applied Meteorology and Climatology*, 52(3), 552–569.
- Lopez, A., de Perez, E. C., Bazo, J., Suarez, P., van den Hurk, B., & van Aalst, M. (2018). Bridging forecast verification and humanitarian decisions: A valuation approach for setting up action-oriented early warnings. *Weather and Climate Extremes*.
- Masutomi, Y., Iizumi, T., Takahashi, K., Yokozawa, M., 2012. Estimation of the damage area due to tropical cyclones using fragility curves for paddy rice in Japan. *Environmental Research Letters*, 7.1.
- Mechler, R., 2005. Cost-benefit Analysis of Natural Disaster Risk Management in Developing Countries. Deutsche Gesellschaft fuer Technische Zusammenarbeit.
- NDRRMC, Memorandum order no. 60, s2019 signed June 2019, Revised guidelines for the declaration of a state of calamity.
- Nobre, G. G., Davenport, F., Bischiniotis, K., Veldkamp, T., Jongman, B., Funk, C., Ward, P., Aerts, J., 2019. Financing agricultural drought risk through ex-ante cash transfers. *Science of the Total Environment*, 653.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pradhan, B., Jebur, M. N., 2017. Spatial prediction of landslide-prone areas through k-nearest neighbor algorithm and logistic regression model using high resolution airborne laser scanning data. Springer, Cham, Switzerland.
- PRISM, 2019 <https://prism.philrice.gov.ph/dynamic-reports/> DynamicReports.
- Probst, P., Wright, M. N., Boulesteix, A. L., 2018. Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, e1301.
- QGIS, 2019 QGIS Geographic Information System Open Source, Geospatial Foundation Project, <http://qgis.osgeo.org>.
- Rajasekaran, S., Gayathri, S., Lee, T.-L., 2008. Support vector regression methodology for storm surge predictions. *Ocean Engineering*, 35.16, 1578–1587.
- Ross, B. C., 2014. Mutual information between discrete and continuous data sets. *PloS one*, 9–2.
- Smola, A., Scholkopf, B., 2004. A tutorial on support vector regression. *Statistics and computing*, 14–3, 199–222.
- USDA, 2017, <https://ipad.fas.usda.gov/highlights/2017/02/philippines/index.htm>
- van den Homberg, M., Monné, R., & Spruit, M. (2018). Bridging the information gap of disaster responders by optimizing data selection using cost and quality. *Computers & Geosciences, Special Issue on Big data and natural disasters: New approaches for spatial and temporal massive data analysis*, 120, 60–72.
- van den Homberg, M., Visser, J., van der Veen, M., (2017) Unpacking Data Preparedness from a humanitarian prioritization perspective: towards an assessment framework at subnational level, *Proceedings of the 14th ISCRAM Conference – Albi, France*, Tina Comes, Frédéric Bénaben, Chihab Hanachi, Matthieu Lauras, eds.
- Van Lint, S., 2016, Sense-making of the Netherlands Red Cross priority index model, case typhoon Haiyan Philippine, <http://edepot.wur.nl/421814>
- Wagenaar, D., Hermawan, T., van den Homberg, M.J.C., Aerts, J.C.J.H., Kreibich, H., de Moel, H., Bouwer, L.M., Improved transferability of multi-variable damage models through sample selection bias correction, *Risk Analysis*, to be published.
- Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., Bai, X., 2015. Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, 527, 1130 - 1141.
- WHO, W. H. O., 2019. Natural Hazards. https://www.who.int/environmental_health_emergencies/natural_events/en
- Wilkinson, E., Weingärtner, L., Choularton, R., Bailey, M., Todd, M., Kniveton, D., & Cabot Venton, C. (2018). Forecasting hazards, averting disasters: Implementing forecast-based early action at scale. *Overseas Development Institute (ODI)*.
- Yang, M. Y. C., Li, Y., 2018. Big data in natural disaster management: a review. *Geosciences*, 8.5