Informatica en Economie

Lowering the Resolution of Damage Data
in a Model that Predicts Rainwater Damage

Teun de Mast

Supervisors:
J.N. van Rijn & M. Baratchi
External Supervisor:
T. Beenen (Stichting RIONED)

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

20/08/2022

Abstract

The increasing risk that residences sustain rainwater damage and the major consequences that rainwater damage can have, highlight the importance of preventive measures being taken. An important step in prevention is finding out which locations are most vulnerable to severe rainfall and hence require more preventive measures. One way to determine this is by developing models that predict rainwater damage for a given amount of rain. There are a variety of data sources that can be used in these models. In this study, we investigate to which degree lowering the spatial resolution of water damage instances that a damage data source provides, is a disadvantage. For this, we look at the performance of three models; one that uses object level water damage instances, one that uses sub-district level instances, and one that uses district level instances. For each resolution, random forest classifiers are used to make predictions for test data. The predictive features that were included in these models are one rainfall feature and multiple height features. A model that considers only the rainfall feature has an accuracy of almost 58%, while a model that considers both rainfall and height features has an accuracy of almost 64%. With the features used in this study, we find small negligible differences in the performances of models considering different resolutions of water damage data. This potentially opens the door to use large insurance data sets, having lower resolutions, to predict rainwater damage.

Contents

1 Introduction

1.1 Background

Together with wind storms, flooding is the most common natural disaster in Europe, accounting for about one-third of the economic losses as a result of natural disasters (Moel and Aerts 2010). There are three main types of flooding:

1. *Fluvial Floods*: Fluvial flooding happens when a river is overflowing due to prolonged periods of heavy rainfall. It can also be brought on by ice jams and heavy snow melt (Maddox 2014);

2. *Pluvial Floods*: During periods of heavy rain, properties may sustain damage from direct roof leakage. When torrential rains produce a flood event independent of an overflowing water body, the result is a pluvial flood (Maddox 2014);

3. *Coastal Floods*: As the name implies, a coastal flood happens in locations that are along the shore of a sea, an ocean, or another significant body of open water. Usually, it is the result of extremely high tides caused by bad weather.(Maddox 2014).

In recent years, extreme cloudburst occurrences that resulted in pluvial flooding have occurred often in Europe. For instance, a 2008 cloudburst in Dortmund caused flooding that cost an estimated 17.2 million Euros to repair (Rözer et al. 2016). In the month June of 2011 a cloudburst hit Copenhagen, resulting in 90,000 insurance claims. With a total cost of approximately 800 million Euros, this became the most costly single incident in Denmark (The Danish Emergency Management Agency 2013). It is commonly believed that pluvial flooding just causes a nuisance with minor consequences. But there are numerous instances where it has led to exposure to contaminants and pathogens, direct loss of life and extensive disruption of transportation networks or other vital urban systems (Rosenzweig et al. 2018).

Urban regions with restricted terrain and coarse soil textures are shown to be more susceptible to pluvial flooding because these locations have significant soil sealing effects (Skougaard Kaspersen et al. 2017). The expected population of the Netherlands of 20 million in 2063 (Centraal Bureau voor de Statistiek 2020), is likely to be coincided with further urban development in the Netherlands. Therefore, if no precautions are taken, it is likely that the Netherlands will be more susceptible to pluvial flooding in the future. For most places worldwide, including Europe, climate change is anticipated to increase extreme precipitation events in both intensity and frequency over the long run (Sunyer et al. 2014). This increases the flooding risk in urban areas (Wheater and Evans 2009) and thus can further increase the risk of pluvial flooding, also in the Netherlands.

The increasing risk of pluvial flooding and the major consequences that pluvial flooding can have, highlight the importance of preventive measures being taken. An important step in prevention is finding out which locations are most vulnerable to severe rainfall and hence require more preventive measures. One way to find this out is by developing models that predict rainwater damage. Machine learning models can be used for this. These models require historical instances of rainwater damage as train and test data. Current models use Twitter messages, P2000 notifications and insurance claims (Lamers, Rijn, and Beenen 2020; Bavelaar et al. 2021; Spekkers et al. 2014). The P2000 network is used by emergency services in the Netherlands to communicate. The instances that these sources provide, differ in spatial resolution. For example, insurance claims are on a district-level while P2000 network notifications can also be on object-level. In this thesis, we research the impact of a lower spatial resolution on a model's ability to predict rainwater damage. We are going to develop multiple models that use water damage instances at different spatial resolutions and analyze their performances. All of these models use a rainfall feature and multiple height features as predictive features. We also look at the effect on the performance of a model when only a subset of the predictive features is used. We expect that a model that uses water damage instances at a low spatial resolution, will perform substantially worse than a model that uses water damage instances at a high spatial resolution. This is because water damage instances at a low spatial resolution can result in information loss in training data.

1.2 Research Question

A strong cloudburst can usually be predicted a few hours in advance, but we know little about what level of precipitation and what aspects of the terrain make a location susceptible to rainwater damage. In this thesis, machine learning methods are applied to predict rainwater damage for a given address. This is only done for addresses in the Netherlands. The features which are designed in this thesis, are derived from a variety of open-source data sources. Historical instances of residences that sustained rainwater damage are used as training and test data. There are multiple data sources available that can provide these instances. The resolution of the instances given by each of these data sources varies; one gives the precise address of the residence that had rainwater damage, while the other just gives the district where the residence that had rainwater damage is located. Low-resolution damage sources might offer advantages like a greater size or more details on the amount of rainwater damage. Consequently, the research question of this thesis is as follows:

> *RQ: What is the effect of lowering the resolution of damage data on the performance of a model that predicts rainwater damage?*

1.3 Contributions

In this thesis, we look at the use of two sources of damage data to predict rain damage; the P2000 network and an insurance data set. We replicate the method to download P2000 network notifications that was used by Bavelaar et al. (2021). We combine feature engineering methods designed by Simons (2021) and the sampling method that was designed by Bavelaar et al. (2021). We create three models that use water damage instances at different levels of spatial resolution and analyze their performances. We also create a model that uses insurance data to predict rain damage. For all of these model we look at the performance of these models when subsets of predictive features are included. Lastly, we analyze a model that uses the predictive feature 'construction year' to predict rain damage (see Appendix A). Code that is developed in this thesis can be found on https://github.com/teundemast/regenwater_overlast.

2 Related Work

2.1 General Trends in Machine Learning

Artificial intelligence finds its origin in 1950 (Turing 2009). Machine learning is a part of artificial intelligence in which a system is trained by providing instances of appropriate input-output behavior without programming the system manually. Over the past two decades, machine learning has advanced significantly, from an idea to a useful tool with extensive commercial use. Machine learning has become the approach of choice in artificial intelligence for creating useful software for computer vision, natural language processing, and other applications (Jordan and Mitchell 2015). The success of machine learning is even presented in famous journals like Science and Nature (Jordan and Mitchell 2015; LeCun, Bengio, and Hinton 2015). However, the biggest issue in machine learning that most algorithms result in a black box system (Holzinger et al. 2018). This is a system that hides its internal logic to the user (Guidotti et al. 2018). These black boxes make judgments non-transparent and difficult to comprehend, which lowers faith in machine learning specifically and artificial intelligence broadly (Holzinger et al. 2018).

A random forest algorithm is a commonly used algorithm in machine learning as it has a high-prediction accuracy and provides information on importance of predictive features (Touw et al. 2012). Although, a system that uses a random forest algorithm is recognized as a black box system, information on the importance of predictive features creates the ability to interpret the system to some extent (Fabris et al. 2018).

2.2 Residential Flood Damage Functions

The application of residential flood damage functions (RFDF) is a frequently used method to evaluate flood exposure and vulnerability to residential areas (Kreibich and Thieken 2008). RFDFs can describe an object's susceptibility to flooding in relative terms (e.g., percentage of total value) or absolute values (e.g., the actual replacement cost of losses). An example of an RFDF is a function which has flood and object characteristics (e.g., the construction material of an object) as input and as output the total amount of damage for this object. Due to complex sewer networks and urban contexts, RFDFs specifically designed for pluvial flood damage have not been designed (Grahn and Nyberg 2017). Besides, regression-based RFDFs are considered unreliable for making predictions due to their limited explanatory power (Grahn 2020). Due to its ability to model complex non-linear relationships, applying a machine learning method might be more suited in predicting floods.

2.3 Predicting Fluvial Flooding from Local Features

The European Commission has made a fluvial flooding simulation available using hydrodynamics of rivers. In 2016, Dottori et al. (2016) used this simulation to produce global hazard maps of return periods for fluvial flooding . A 'flood return period' is the projected time elapsed between floods of comparable size or intensity (*What is a return period?* 2014). The method used by Dottori et al. (2016) has two downsides: the lack of explainability and the lack of adaptability. Both of these are important factors for the resilience of a flooding prediction model. To address these issues, Ronk, Splunter, and Knibbe (2022) developed 31 local features in an attempt to predict fluvial flooding. In this context, local features are those that arise from the location of an object itself or its near surroundings. These features were mostly continuous variables. The features that Ronk, Splunter, and Knibbe (2022) developed, were grouped into three categories based on their core 'area of effect':

1. *Micro*: features that were created using data that encompasses elements on the exact location. An example is the one day precipitation maximum;

2. *Meso*: features that were created using information from nearby surroundings. An example is the shortest distance to a river;

3. *Macro*: features that encompass information about a larger area of effect. An example is the long term vision of the government.

Ronk, Splunter, and Knibbe (2022) compared three machine learning methods that used these features to solve the binary task: will a location experience flooding in the next 20 years or not? The classifiers that were used to solve this task are: a logistic classifier, a random forest classifier and neural networks. The results showed that the predictions made by the random forest classifier had the highest accuracy. Ronk, Splunter, and Knibbe (2022) concluded that local features can be used to assess a risk of fluvial flooding with sufficient accuracy. According to the results, micro features describing ground imperviousness (the incapability of ground being penetrated by water) contributed significantly to the predictive power of the model. Also, the meso features about 'relative height' played a major role in the prediction of fluvial floods. This type of features provides information about the relative height of an object to its surroundings (Ronk, Splunter, and Knibbe 2022).

2.4 Predicting Rainwater Damage from Local Features

In 2019, Bernet et al. (2019) conducted research on the precipitation patterns that were most likely to result in damage-relevant pluvial floods. Insurance claims in Switzerland were used to perform the analysis. According to their analysis of damage-relevant and non-relevant precipitation events, the features precipitation intensity and total precipitation discriminated most (Bernet et al. 2019). As shown in Figure 1, events with a high intensity and high total sum are more likely to result in more insurance claims in Switzerland. Therefore it is likely that precipitation with a high intensity and a high sum are more likely to cause damage-relevant pluvial floods.

In 2014, Spekkers et al. (2014) investigated a wide range of features associated with rainfall-related damage using decision-tree analysis. For this, district-aggregated claim data from Dutch private property insurance firms were analyzed. According to the study, claim frequency is correlated with real estate value, building age, maximum hourly rainfall intensity and other local features. A large part of the variance was left unexplained, which was likely to be caused by variations in data at sub-district scale and missing explanatory variables (Spekkers et al. 2014).

Instead of insurance data, Lamers, Rijn, and Beenen (2020) used Twitter messages referencing rainwater damage as a proxy variable for rainwater damage. In this study, a height map and an amount of precipitation were used as local features used to predict rainwater damage (Lamers, Rijn,

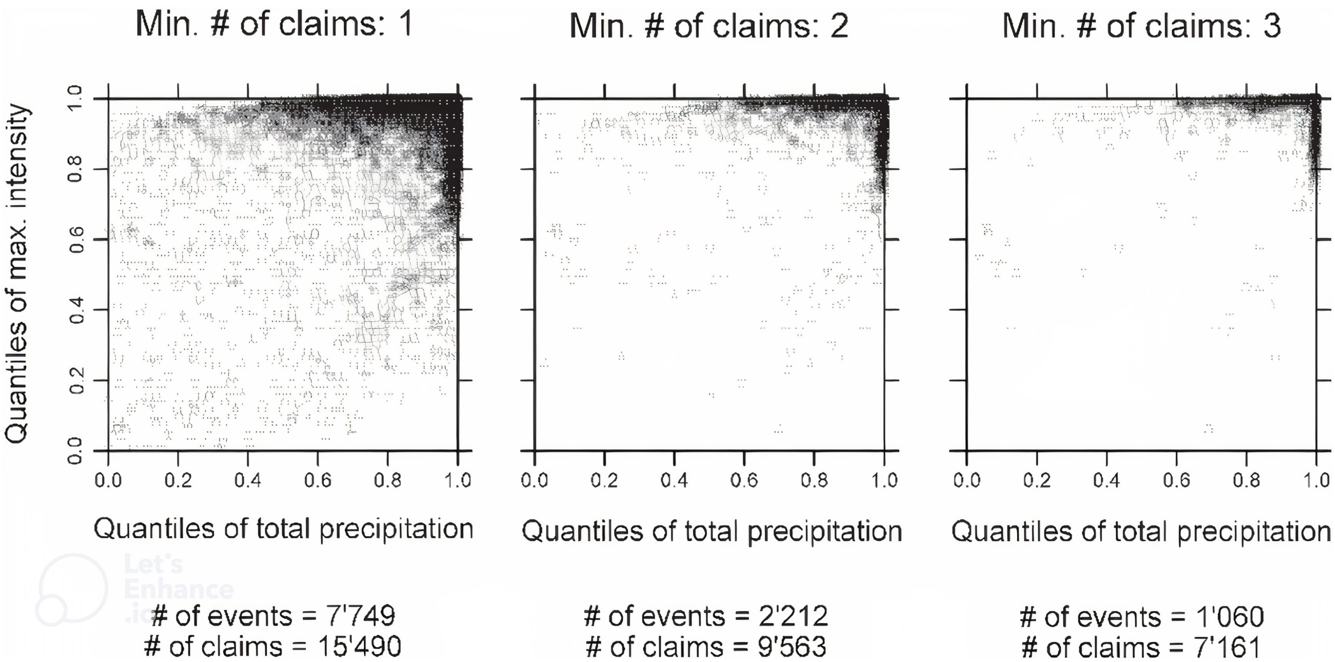

Figure 1: The effect of total precipitation and rain intensity on the number of insurance claims related to pluvial floods (Bernet et al. 2019). As shown in the most right figure, (almost) only pluvial floods with a high total precipitation and a high intensity, will result in three or more insurance claims.

and Beenen 2020). In 2021, Bavelaar et al. (2021) proceeded using this model and compared the use of Twitter messages to the use of P2000 messages. P2000 messages are messages to emergency services. Bavelaar et al. (2021) concluded that P2000 messages are superior to Twitter tweets because they result in less noise in the data. Since these messages only contain positive instances of rainwater damage, negative instances need to be sampled. Bavelaar et al. (2021) compared the following three sampling methods to do this:

1. *Random*: The negative instances have a height map of a random location in the Netherlands. They also have a random amount of precipitation above some threshold;
2. *Address*: The negative instances have a height map of a random address in the Netherlands. They also have a random amount of precipitation above some threshold;
3. *Dependent*: The negative instances have the same height map as the positive instances but have a random amount of precipitation above some threshold. This model aims to answer the question: "Why do areas sometimes experience damage from rainfall and sometimes not?" Intuitively, rain could damage both lower and higher located residences, but highly situated areas might require a higher amount of precipitation to encounter problems.

Bavelaar et al. (2021) concluded that the second and the third sampling method are most suited for predicting whether a residence sustains rainwater damage or not.

The model created by Bavelaar et al. (2021) and their conclusions are used as a starting point for this thesis.

3 Data

In this section we describe the variables and data sources that are used in models in this study. The target variable will be described first, followed by a description of the predictive features that are engineered to predict the target variable.

3.1 Rainwater Damage Variable

The target variable of the models that are created is a binary variable which answers the question: 'Does a residence at location X experience rainwater damage at timestamp Y or not?'. Multiple data sources can provide a proxy-variable for this target variable.

A private insurance party in the Netherlands has provided a data set for this study that includes the number of insurance claims per day per district. The data set includes more than a million water-related damage claims in the period 2017-2022. A claim provides a date on which the water damage has occurred. Due to privacy concerns, the data is aggregated at the level of four-digits postal districts. Exploring a database provided by Statistics Netherlands in 2021, data source 4 in Table 1, resulted in the fact that there are 4,069 four-digits postal codes in the Netherlands. The lowest postal code is 1011 and the highest is 9999. Of the 4069 possible four-digits postal codes, there are more than 99% of the postal codes present in the insurance data set. The district with the most water-related claims in the period 2017-2022, is 2134, a district located in Hoofddorp.

#	Data source	Temporal resolution	Spatial resolution	Period	Note	Publicly available
1a	Database from a private insurance party	By day	District level	2017-2022	Number of claims per district is available	No
1b	P2000 network	Exact time	Per object	Obtained in the period 2016-2021	Notifications related to water damage	Yes
2	Weather data set from the Royal Netherlands Meteorological Institute	Every 5 minutes	1km	Obtained in the period 2016-2021		Yes
3	Digital height map of the Netherlands (AHN3)	1 scan	0.5m x 0.5m pixels	Snapshot	Use has been made of an available API	Yes
4	Database from Statistics Netherlands	1 scan	-	2021	Containing information on which sub-district is part of which district	Yes
5	National Building Register	1 scan	Per object	Snapshot	Use has been made of an available API.	Yes

Table 1: Overview of the databases used in this thesis.

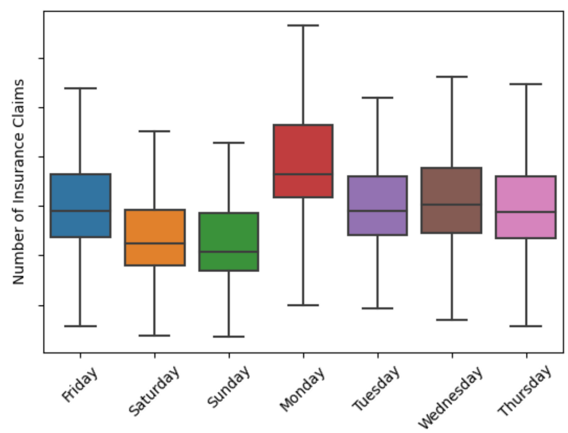

Figure 2: The distribution of insurance claims per weekday. Note: the y-axis does not show values due to privacy reasons.

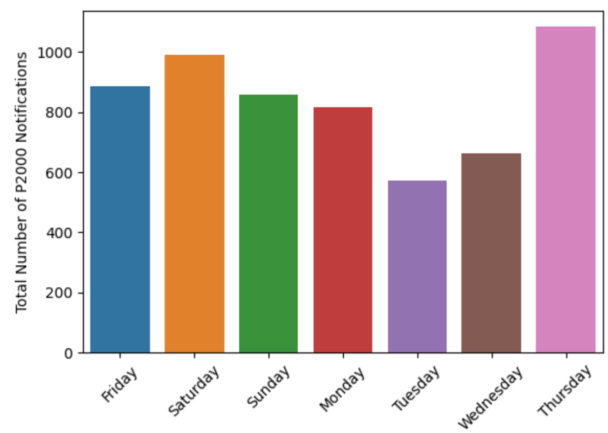

Figure 3: Total number of precise P2000 notifications concerning water damage grouped per weekday.

There is bias present in the insurance data set. Some of the bias finds its origin in the following two phenomena:

- As shown in Figure 2, there are on average less claims in the weekend and more claims on Monday. This can be due to the fact that people who are away from their home over the weekend, notice the water damage the first time on Monday. They then are unable to know if the water damage occurred on the weekend or on Monday.
- Claims that were submitted without an exact date of the occurrence of water damage are automatically set on the first day of the month. This creates an unrealistically high number of insurance claims on the first day of each month.

As opposed to the insurance data set, the P2000 network can be used to generate instances of water damage at an object level resolution. The P2000 network is a network used by emergency services in the Netherlands to communicate. While using an API, notifications in this network are downloaded for the period 2016-2021 applying a method provided by Bavelaar et al. (2021). Only the messages that refer to water nuisance are considered to be valuable. It is assumed that a message relating to water nuisance implies an instance of water damage. Unlike insurance claims, P2000 messages provide an exact timestamp of when a message was sent out. However, a major problem with this data set is that it only provides around 5700 instances of water damage at an object-level. There are more P2000 notifications available when considering notifications at a street, sub-district or district level. However, for these notifications it is also likely that not a residence sustained water damage but a car for example. Therefore, when predicting rainwater damage for residences, these notifications would result in a lot of noise in a data set of instances of rainwater damage. When all of the object-level instances are grouped by day, small differences in frequency between days can be observed (see Figure 3). Due to the small size of the data set, no conclusion could be drawn from these differences. Nevertheless, it is unlikely that the event of a person notifying emergency services due to water-related nuisance is related to a weekday.

Due to the fact that insurance claims are directly related with damage, they might be a better proxy-variable for rainwater damage than P2000 notifications. In addition, there are significantly more water-related insurance claims than water-related P2000 notifications. In 2021, there were around a 1000 water damage related P2000 notifications in the months October, November and December. In comparison, there were at least 30,000 water-related insurance claims in these months. However, all insurance claims are aggregated at a district-level while some of the P2000 notifications are available at object-level. The advantages and disadvantages of insurance claims are the motive for this study, as described in Section 1.2. In this study, three models are created with P2000 notifications at different resolutions (Sections 4.1.1, 4.1.2 and 4.1.3) and one with insurance data (Section 4.1.4).

3.2 Features

In this study, two local features are designed for every instance of rainwater damage. These are features that are based on a previous study on this topic (Bavelaar et al. 2021). The first feature is a micro feature and the second a meso feature (as defined in Section 2.3). The features are defined as follows:

1. *3 hours rainfall*; this variable is equal to the sum of the rainwater that fell in the 3 hours preceding an occurrence of water damage. This information is retrieved using a weather radar data set from the Royal Netherlands Meteorological Institute. The method used for this is provided by Bavelaar et al. (2021). This data set consists of measurements of the amount of rainfall for every five minutes in one kilometer grids. In this study, the amount of rainfall for an address is equal to the amount of rainfall in the 1km grid this address is located in;

2. *Height 1 … 400*; using a digital height map of the Netherlands, a 10m x 10m height map is retrieved for each instance of rainwater damage. The method used for this is provided by Simons (2021). The coordinates of the residence that experienced rainwater damage are in the center. Each 0.5m x 0.5m has one height value, resulting in an array of 400 values. Each value in such an array is seen as a feature. The choice of a 10m x 10m is an arbitrary one. Using a larger height map was considered but the model seemed to perform worse. Larger height maps result in a lot more predictive features and make it hard for a random forest classifier to find informative features. A smaller height map might result in information loss in the data set.

In the Netherlands, there are four different digital height maps: 'AHN1', 'AHN2', 'AHN3' and 'AHN4'. Each differs in resolution and precision.[1] Bavelaar et al. (2021) used 'AHN2' in a model to retrieve height features. In this study, 'AHN3' is used to achieve more precise height features.

1. More information about these differences can be found here: https://www.ahn.nl/kwaliteitsbeschrijving

Variable name	Definition	Source
Target variable		
Rainwater damage	This variable is a binary variable that is predicted by the model. It answers the question: does an address experience rainwater damage? (yes/no). When data source 1a is used, it is assumed that insurance claims imply rainwater damage. If data source 1b is used, it is assumed that a P2000 notification related to 'water damage' and the cause is probably rain, indicates an occurrence of rainwater damage.	1a/1b
Local features		
3 hours rainfall (Micro)	The amount of rainfall that fell on an address preceding an occurrence of water damage.	2
Height 1 ... 400 (Meso)	400 height features, each representing one pixel in a height map. In the center of this height map, the residence that sustained rainwater damage is located.	3
Optional:	(This variable is used in a case study in an appendix)	
Construction year (Micro)	The construction year of a residence that sustained rainwater damage	5

Table 2: Overview of the variables that are used in this thesis.

4 Method

To answer the research question *What is the effect of lowering the resolution of damage data on the performance of a model that predicts rainwater damage?*, three models are created. Each of these models is created using different resolutions of damage data. As is described in Section 3.1, the P2000 network provides water damage instances at an object level. The first model will be created using these precise instances. Another model is created with sub-district level instances of water damage and another with district level instances of water damage. The instances of water damage of these first three models are derived from the object-level instances provided by the P2000 network. The performance of each of these models will be analyzed to draw a conclusion. Keep in mind that the temporal resolution of the water damage instances is the same for of these models.
In this study, a fourth model is created with the use of water damage instances provided by the insurance data set. The temporal resolution of these instances differ from water damage instances provided by the P2000 network.

4.1 Engineering the Data Sets

The first step for each of the models is engineering a data set in which the variables described in Table 2 are present. The process of engineering a final data set is different for each of the models. They are described in the following subsections.

4.1.1 P2000: Object Level Instances

Precise water damage instances provide an exact address that sustained water damage. Also, the exact time of the occurrence of water damage is available. The following steps are taken to generate a data set on which a machine learning method can be applied (see Figure 4):

1. *Enrich with rain information*
   The water damage instances are enriched with rain information, specifically with the amount of rain that fell in the 3 hours preceding the water damage. This creates the variable '3 hours rainfall' as defined in Section 3.2.;

2. *Find rainwater damage instances*
   In this study, only instances of water damage that were caused by rain are relevant. Therefore, instances of water damage that were not caused by rain but by any other source need to be filtered out. The following assumption is made: 'If the amount of rain that fell in the 3 hours preceding an instance of water damage is higher than 50 mm, the cause of the water damage is rain.'. The choice of this threshold is made by Simons (2021) and is somewhat based on the capacity of sewer systems in the Netherlands. However, this threshold could be optimized. Under this assumption, instances of water damage that were not caused by rain are filtered out;

3. *Add height features*
   Using the address provided by each instance of rainwater damage, a 10m x 10m height map is retrieved with the coordinates of this address in the center. Since every value in the height map is a measure for a 0.5m x 0.5m square, the height map consists of 400 values. All of these values are added as features (as described in Section 3.2);

4. *Sample negative instance*
   For each positive instance, one negative instance is added to the final data set. This is done using a 'dependent' sampling method, which is defined in Section 2.4. A negative instance has the same height features as the positive instance it originated from. The '3 hours rainfall' feature for this negative instance is a random number between the threshold of 50mm and the value of the '3 hours rainfall' of the positive instance. It is assumed that for this amount of rainfall in three hours, the address will not sustain any water damage.

4.1.2 P2000: Sub-District Level Instances

In the Netherlands, sub-districts are characterised by a six-digits postal code. Using Statistics Netherlands, data source 4 (see Table 1), it can be found that there are around 460,000 sub-districts in the Netherlands. On average, a sub-district consists of around 17 residences.
Each of the object-level instances used in Section 4.1.1, is transformed in an instance where the address is a random address in the same sub-district (see Figure 4). This is done by using data source 5 (see Table 1), the National Building Register. Subsequently, the water damage instance is considered to provide an address (the generated random address) and the steps described in Section 4.1.1 can be executed.

4.1.3 P2000: District Level Instances

Water damage instances at district-level only provide a four-digits postal code of the residence that sustained water damage. For this case, the precise water damage instances, provided by the P2000 network, are considered to only provide a four-digits postal code of the residence that sustained water damage. To be able to execute the steps denoted in Section 4.1.1, the district level water damage instance needs to be transformed into a precise water damage instance (see Figure 4). First, using data source 4 (see Table 1), a random six-digits postal code is select where the first four digits match the four-digits postal code that is provided by a water damage instance. Subsequently, one residence is randomly selected that has the generated six-digits postal code. This is done by using the National Building Register. Subsequently, the water damage instance is considered to provide a precise address (the generated random address) and the steps described in Section 4.1.1 can be executed.

4.1.4 Insurance: District Level Instances

In this study, the performance of a model that uses insurance claims as water damage instances, is compared to the performance of a model that uses P2000 notifications as water damage examples. The insurance data set provides over one million instances of water damage while the P2000 network only offers around 5700 water damage instances. Taking computing time into consideration, we only used around 15,000 insurance claims as a sample.

To be able to compare the performance of a model that uses insurance claims with a model that uses P2000 notifications, the steps described in Section 4.1.1 need to be executed also for the insurance claims. Since the insurance data set provides water damage instances at a lower temporal resolution (i.e., per day) than the P2000 network (Table 1), steps need to be taken to address this lack of information. To engineer the variable '3 hours rainfall', an exact timestamp of the occurrence of water damage needs to be provided. Therefore, it is assumed that a water damage instance provided by the insurance data set occurred right after the highest sum of three hours of rainfall in the given day.

Subsequently, the steps described in Section 4.1.3 are executed.

4.2 Using a Machine Learning Algorithm to Train a Model

In Section 4.1 is described how four data set are engineered using different resolutions of water damage instances; one at object level, one at sub-district level and two at district level. Since the target variable is a binary variable (does an object experience rainwater damage or not?), a classifier needs to be trained. The best performing classifier for predicting fluvial flooding using local features was a random forest classifier (Ronk, Splunter, and Knibbe 2022). Therefore, this machine learning method was also used in this study. A random forest classifier[2] is trained for each of the above-mentioned types of data set. The hyperparamaters that are are noted in Table 3.
In the following paragraph is described how training and test data sets were created.

2. The random forest classifier from scikit-learn version 1.0.2 is used in this study (Pedregosa et al. 2011)

Hyperparamater	Value
Number of trees in the forest	1000
The function to measure the quality of a split	gini
Minimum number of samples required to split an internal node	2
The minimum number of samples required to be at a leaf node	1
The number of features to consider when looking for the best split	square root of number of features

Table 3: The hyperparameters used in this study for random forest classifiers.

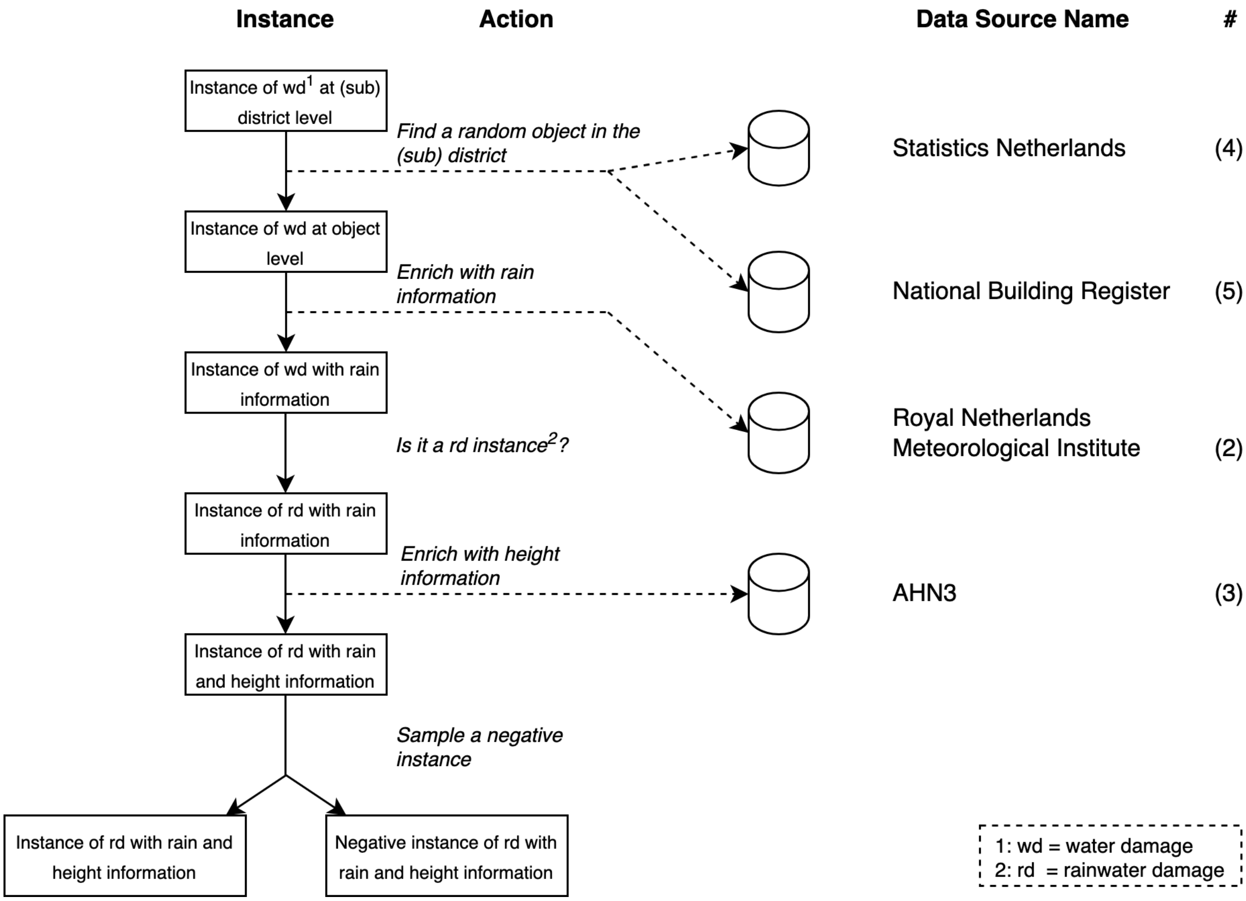

Figure 4: The pipeline that is used to engineer data sets. If a water damage is already at object level, the first action is not executed.

4.3 Training and Testing

The final pipeline that is used to engineer data sets, train classifiers and test the performance of classifiers, is visualized in Figure 5. The following section focuses on how training and test data sets were created from the engineered datasets in Section 4.1.

In the process of engineering data sets there are multiple steps that include randomization. For engineering a data set that uses object-level instances of water damage, the only step that includes randomization is '*Sample negative instance*'; for a negative instance, the value for the variable '3 hours rainfall' is chosen randomly. When dealing with sub-district or district level instances, the precise address is also chosen randomly. To make sure the results that are found when testing the models are not coincidental, five data sets were generated for each of the resolutions. Each of the five data sets that was generated using object-level instances, was split up into ten smaller data sets that could take the role of test set. This resulted in a total of fifty possible test data sets. Since the data sets that were engineered for sub-district and district level instances, contain generated bias, data in these sets can not be used as test data.

For each of the five data sets per resolution, random forest classifiers were trained and tested considering ten random test data sets of the fifty test data sets available. During this process, it was made sure that no instances were present in the training ánd the test set. Therefore, the instances in a test set where considered before determining which instances should be present in the training data. This meant that for each of the five generated data sets per resolution, ten random forest classifiers were trained and tested. Thus, for each resolution, fifty random forest classifiers were trained and tested. Also, each training set was set to include 1700 instances of rainwater damage. However, we were only able to locate 700 instances of rain damage in the data set that was constructed using insurance claims. During the testing process, three performance measures were used; accuracy, precision and recall. They are defined as followed:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \tag{1}$$

$$Precision = \frac{True\ Positives}{True\ Positives\ + False\ Positives} \tag{2}$$

$$Recall = \frac{True\ Positives}{True\ Positives\ + False\ Negatives} \tag{3}$$

Precision, as defined in Equation 2, attempts to answer the question: "What proportion of positive identifications was actually correct?". Recall, as defined in Equation 3, tries to answer the following question: "What proportion of actual positives was identified correctly by the classifier?" (*Classification: Precision and Recall*).

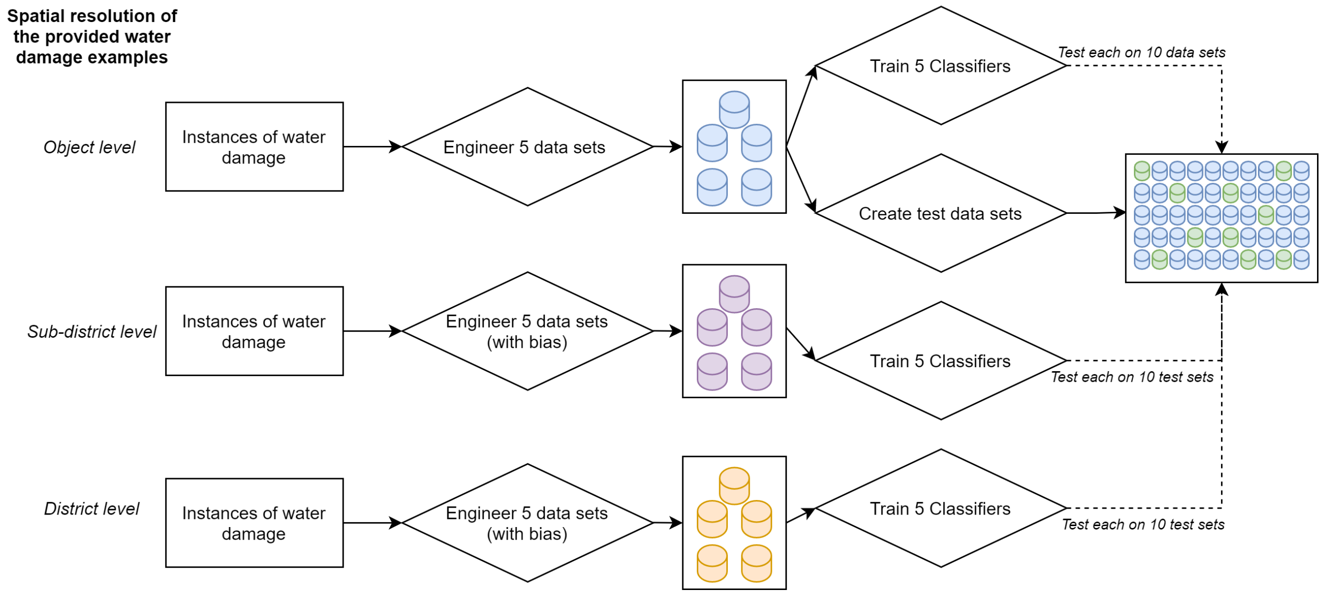

Figure 5: The pipeline that is used to engineer data sets, train classifiers and test classifiers.

5 Results

In this section the results are given for classifiers that predict if an object sustains rainwater damage based on two kind of local features: rainfall and height. To determine the importance of considering both of these features, three kind of classifiers were made for each resolution; one which only considers the rainfall feature, one which considers only height features and one which considers both. The performance of each of these classifiers is described in the following subsections. Results of classifiers that were developed using P2000 network notifications are given in the first three sections. Results of classifiers that were developed using insurance claims are given in the fourth section.

5.1 Predictions Made Considering One Rainfall Feature

Spatial resolution of water damage instances	Average accuracy	Average precision	Average recall
Object level	0.5758	0.5790	0.5635
Sub-district level (w.r.t. object level)	+ 0.0083	+ 0.0116	+ 0.0112
District level (w.r.t. object level)	+ 0.0064	+ 0.0084	+ 0.0161

Table 4: The performance on test data sets (N= 50) of classifiers that consider different spatial resolutions of water damage instances. They only consider one rainfall feature.

In Table 4 and Figure 6 the accuracy, precision and recall scores can be found of a model that makes predictions considering only the '3 hours rainfall' feature as defined in Section 3.2. The results show that the models perform better than a model that classifies rainwater damage instances randomly

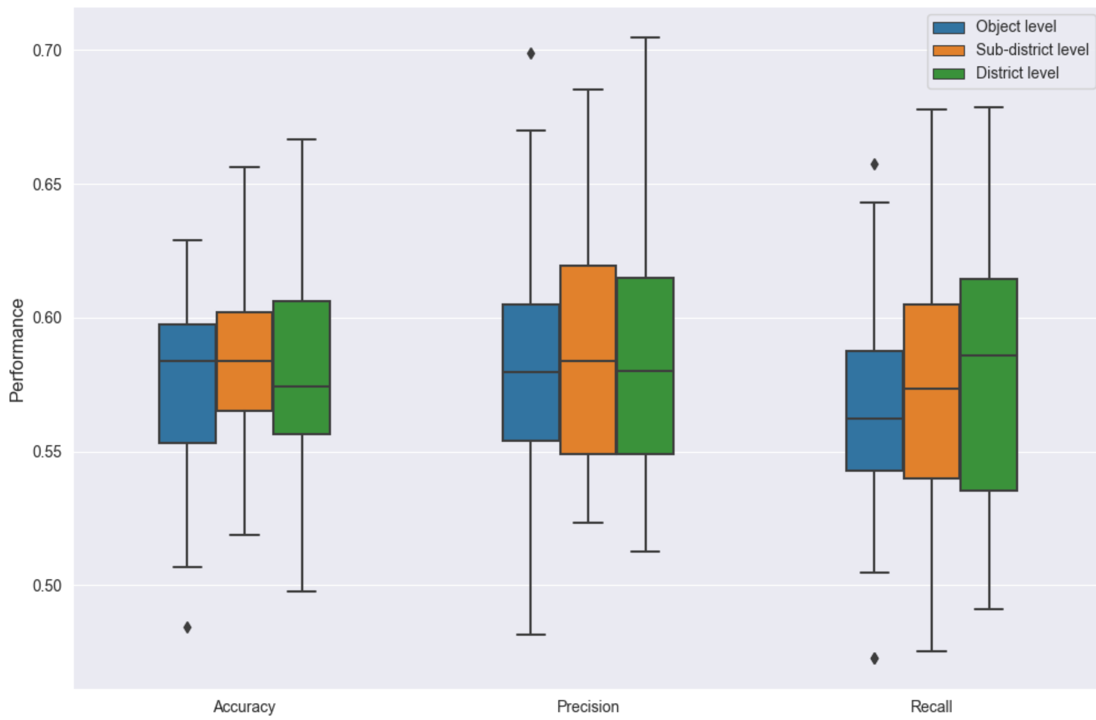

Figure 6: A box-plot with the performance of classifiers that consider different spatial resolution of water damage instances. They can only include the rainfall feature.

as they have a higher accuracy than 50%. The differences in average performance between the models using different resolutions of water damage instances, are less 2% (Table 4).

5.2 Predictions Made Considering Height Features

Spatial resolution of water damage instances	Average accuracy	Average precision	Average recall
Object level	0.4974	0.4987	0.4867
Sub-district level (w.r.t. object level)	+ 0.0019	+ 0.0029	+ 0.0040
District level (w.r.t. object level)	+ 0.0033	- 0.0001	+ 0.0162

Table 5: The performance on test data sets (N= 50) of classifiers that consider different spatial resolution of water damage instances. They can only include height features.

In Table 5 and Figure 7 the accuracy, precision and recall scores can be found of classifiers that make predictions considering only height features. There are 400 height features as described in Section 3.2. The results show that the classifiers perform almost as good as a classifier that classifies rainwater damage instances randomly.

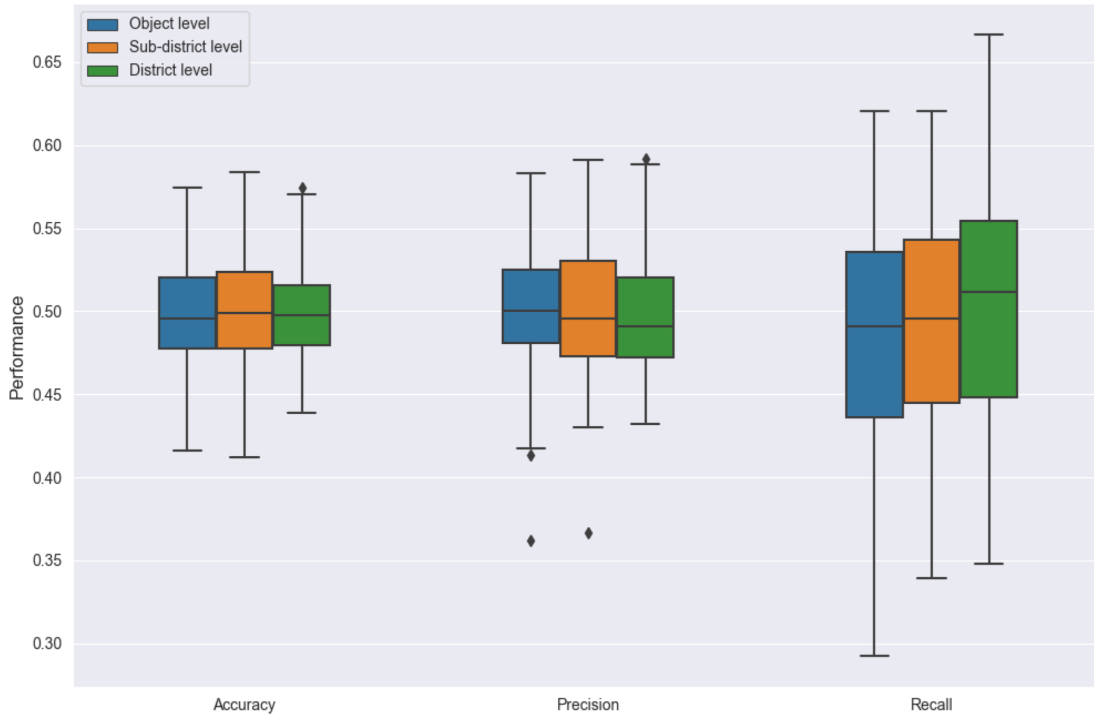

Figure 7: A box-plot with the performance of classifiers that consider different spatial resolution of water damage instances. They can only consider height features.

5.3 Predictions Made Considering Both Rainfall and Height Features

Spatial resolution of water damage instances	Average accuracy	Average precision	Average recall
Object level	0.6367	0.6418	0.6280
Sub-district level (w.r.t. object level)	- 0.0027	- 0.0031	- 0.0053
District level (w.r.t. object level)	- 0.0016	- 0.0107	+ 0.0071

Table 6: The performance on test data sets (N= 50) of classifiers that consider different spatial resolution of water damage instances. They can include one rain feature and height features.

In Table 6 and Figure 8 the accuracy, precision and recall scores can be found of classifiers that make predictions considering height features and one rain feature. The results show that the classifiers perform better than a classifier that classifies rainwater damage instances randomly. It also performs better than a model that considers only a rain feature or only height features.

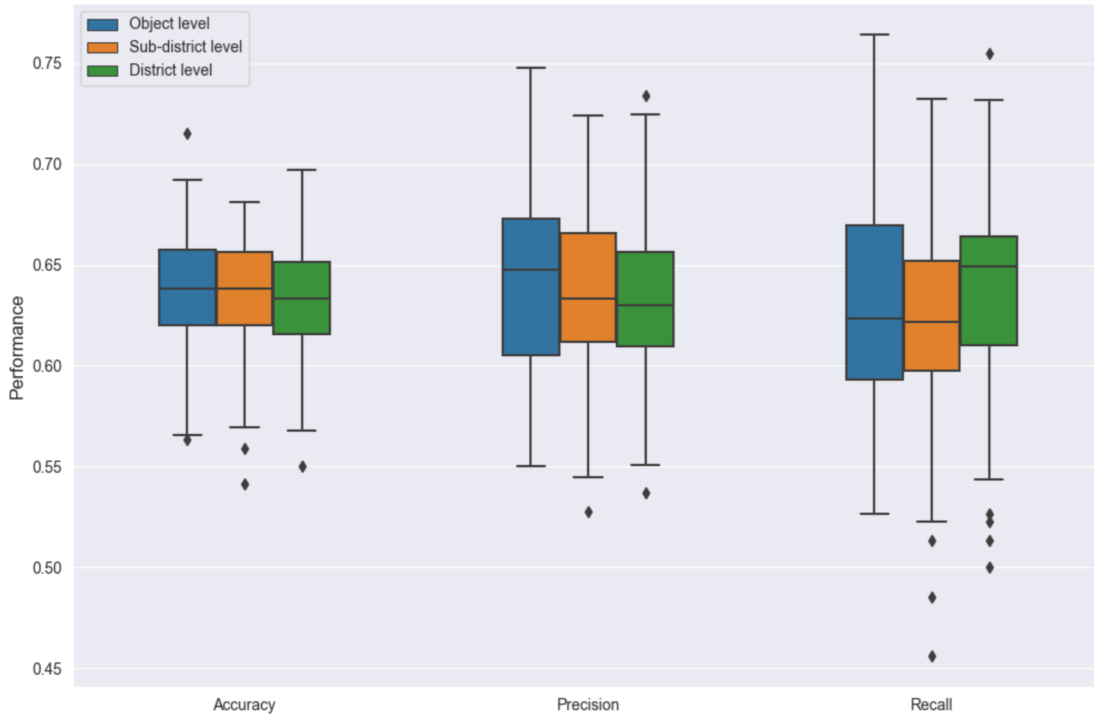

Figure 8: A box-plot with the performance of classifiers that consider different spatial resolution of water damage instances. They can include one rain feature and height features.

5.4 Prediction Made Using Insurance Claims

In Table 7 and Figure 9 the accuracy, precision and recall scores can be found for classifiers that make predictions considering different subsets of predictive features. Especially for the evaluation measure recall, the difference in performance between these classifiers is the largest.

Included predictive features	Average accuracy	Average precision	Average recall
Only rain features	0.5748	0.5561	0.7569
Only height features	0.5014	0.4990	0.5407
Both rain and height features	0.5973	0.5634	0.8838

Table 7: The performance on test data sets ($N = 50$) of classifiers that consider different subsets of predictive features. The classifiers were trained using insurance claims.

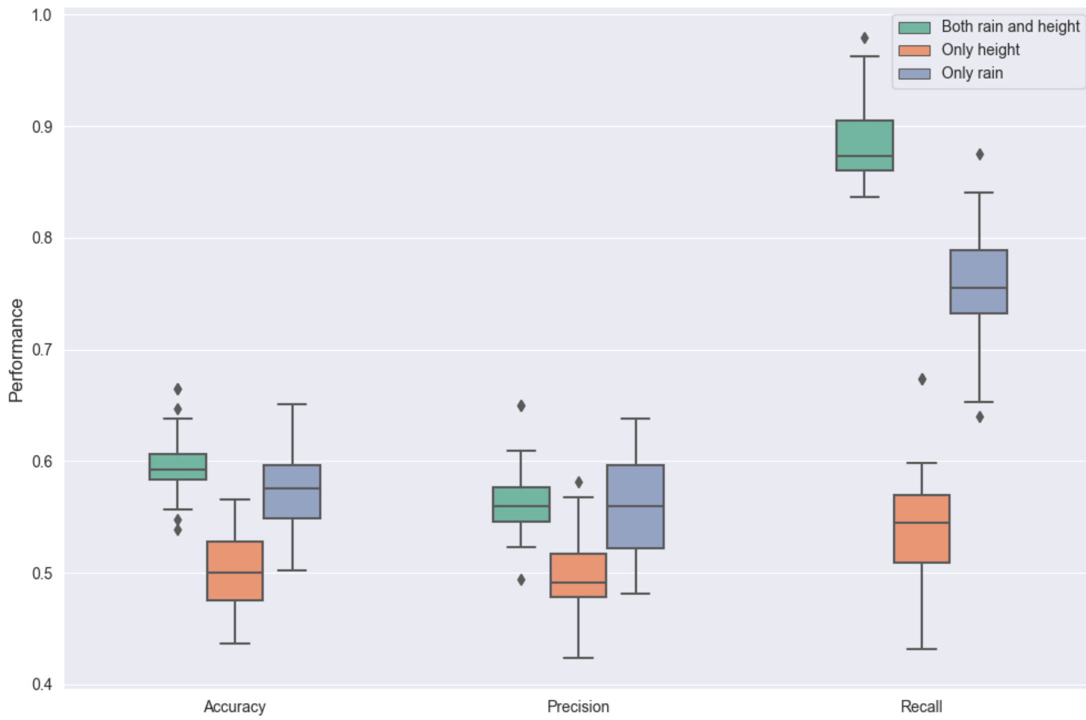

Figure 9: A box-plot with the performance of classifiers using different subsets of predictive features. The classifiers were trained using insurance claims.

6 Discussion

We examined the performance of classifiers differing in spatial resolution of the used information (object-level, sub-district level and district-level) and differing in types of input features (only rainfall, only height variables, or both). The classifiers were tested using the same object-level test data sets. With regard to the results for the different resolutions in the P2000 network data, the differences in average accuracy, precision, and recall between the classifiers derived from them, were not larger than 1.7% and often close to 0%. With regard to the results using different input features, we found the best performance for those classifiers using both the rainfall feature and the height features.

Although no statistical tests were performed, it is unlikely that any of the found differences in performance of the classifiers using different resolutions is statistical significant. Even if they are, the differences are not practically significant, meaning that choosing between sources of water damage instances, should probably not depend on the spatial resolution of those water damage instances. Note that this is just the case for predicting rainwater damage when only considering the height features and rain feature defined in this study. In addition, it is only the case when the same sampling method is used as in this study.

A possible explanation for these findings is that height maps of residences in the same district are probably not that different. This results in the fact that the random forest classifiers in this study are somewhat able to determine which kind of districts (based on height maps) are more susceptible to rainwater damage if a particular amount of rain falls. However, the classifier is probably unable to determine which residences in a district are more susceptible to rainwater damage compared to other residences in its district. There are often cases of heavy rain where only one or two residences in a district experience rainwater damage. To determine these residences, a classifier will probably need to be able to use more object-level features. If more object-level features were added to a model, the resolution of water damage instances is likely to matter a lot.
As stated in Section 5.2, a model that predicts rainwater damage using only height features does not perform better than a model that classifies randomly. This is because positive and negative instances that are derived from the same rainwater damage instance, have the same height map as described in Section 4.1. Therefore, a classifier can not use these features to make a better decision than 'flipping a coin'.

Our results indicate that a model that predicts rainwater damage using both a rainfall feature and height features, is likely to perform considerably better than a model that only uses a rainfall feature. This might not be intuitive because a classifier that uses only height features as predictors predicts as good as random. However, it is likely that whenever predictions are made with both height features and the rainfall feature, 'interaction effects' are present. This means that the effect of one variable depends on another and therefore synergistic information is available (Quax, Har-Shemesh, and Sloot 2017). In this case, the effect of height features is dependent on rain information. It is therefore likely that a particular amount of rainfall will impact some objects more than others due to different height patterns.

The classifiers that were trained using insurance claims had a considerably lower accuracy and precision than classifiers that were trained using P2000 notifications at the same spatial resolution (i.e., district level). However, they had a substantially higher recall score. Note that all classifiers were tested on test data that was generated using object level P2000 notifications. We find that the value of the variable '3 hours rainfall' is often much higher for rain damage instances that were derived from P2000 notifications than for rain damage instances that were derived from insurance claims. This is probably due to the fact that a person only notifies emergency services if the person is in danger. This will most likely happen when it rains heavily. During periods of relatively little rain, there is a higher chance that a person does file an insurance claim but does not notify emergency services. As a result, classifiers trained on insurance data will predict a positive instance more frequently than classifiers trained on P2000 notifications.

7 Limitations

This study has four main limitations. One limitation is that a random forest solution is not a fully explainable solution. It is possible to analyze some trees in a forest but not all of them. Besides, the height features in this study are pixel-based and therefore splits in the tree on the value of these pixels are hard to comprehend. This might hamper the resilience of the prediction model. A single decision tree has been used to analyze decisions made in a model that predicts rainwater damage (Spekkers et al. 2014). However, a random forest is likely to perform better for larger data sets than a single decision tree (Ali et al. 2012).

Another limitation is that rainfall information for negative instances is generated randomly. An assumption was made that for this amount for rain, no rainwater damage would occur. If this assumption does not hold for all examples, which it probably does not, it can lead to false negative instances.

Also, the height map that is retrieved for each instance, as described in Section 4.1.1, is only a 10m x 10m height map. This might mean that relevant height measure outside of the retrieved map are left out. Therefore, using a larger height map can result in more information being available in the features. We did do a few experiments with models that use larger height maps, which did not improve the performance of a model. However, this may be due to the fact that then too many pixel-based height features are included in the model, which results in high-dimensional data. The performance of a random forest algorithm is usually seriously weakened when facing high-dimensional data as it can contain irrelevant and redundant features (Feng et al. 2020).

The National Building Register was accessed using an API which is limited in the number of requests. In the process of engineering data sets using (sub) district level water damage instances, the National Building Register needs to be accessed for each instance (Section 4.1). This resulted in the fact that only around 15,000 insurance claims were used to minimize computation time. In the process of engineering data sets using this sample (Section 4.1.4), it was often concluded that the insurance claims were not filed because of rain damage. As a result, the engineered data sets only contained around 700 rain damage instances. The size of a training set that was created using insurance samples was not even half the size of a training set that was created using P2000 notifications. As classifiers tend to perform better when a larger training set is used, our comparisons between classifiers that use P2000 notifications and classifiers that use insurance claims might be unfair.

8 Conclusions and Further Research

There are a variety of data sources that can be used as a damage data source in a machine learning model that predicts if a residence sustains rainwater damage or not. In this study, we researched the effect of a low spatial resolution on the performance of a model. For this, we have looked at the performance of three models that use P2000 notifications; one which uses object level water damage instances, another which uses sub-district level instances and one that uses district level instances. Another model was created using insurance claims. A random forest classifier was used as machine learning method to make predictions for test data. The predictive features that were included in these models were one rainfall feature and multiple height features. The differences in accuracy across classifiers that use P2000 notifications at different level of spatial resolutions, were smaller than 1%. It can be concluded that for the created models, the spatial resolution of data has negligible impact on the performance. However, the models that were used in this study did not include many local features, such as the building material and the total value of a residence. When more local features are added, the spatial resolution of water damage examples might be important.

Furthermore, it can be concluded that random forest classifiers that include both rain and height features perform better than classifiers that include only height features or only a rain feature. This is the case for models that use P2000 notifications and models that use insurance claims.

We noted that the insurance data have a lower temporal resolution than P2000 notification. In the future, the impact of the temporal resolution of water damage examples on the performance of models that predict rainwater damage, can be investigated. As described in Section 7, only a sample of the entire insurance data set was used. By using the whole insurance data set in future research, classifiers might perform better. In Section 7 is also described that the size of height maps that were used, might not be optimal. Using AutoML, the size of these maps can be optimized (Karmaker ("Santu") et al. 2021). Future research may also include more features to improve predictive performance and to enable the efficient use of measures to prevent rainwater damage. For example, the local features that were engineered by Ronk, Splunter, and Knibbe (2022) to predict fluvial flooding, can also be implemented to predict rainwater damage. Furthermore, the performance of different type of machine learning methods can be analyzed. Deep neural networks might outperform random forests. Decisions that deep neural networks make, are frequently seen as being difficult to understand, hence they might not be appropriate for predicting rain damage. Recent research, however, has proposed explainable deep neural networks that could resolve this problem (Angelov and Soares 2020; Vaughan et al. 2018).

References

Ali, Jehad, Rehanullah Khan, Nasir Ahmad, and Imran Maqsood. 2012. "Random Forests and Decision Trees." *International Journal of Computer Science Issues,* 3rd ser., 9 (5): 272–278.

Angelov, Plamen, and Eduardo Soares. 2020. "Towards explainable deep neural networks (xDNN)." *Neural Networks* 130:185–194.

Bavelaar, Christie, Thijs Simons, Jan van Rijn, Mitra Baratchi, and Ton Beenen. 2021. "Toepassing van machine learning-technieken op voorspelling van regenwateroverlast in stedelijk gebied." *H20.*

Bernet, Daniel Benjamin, Simona Trefalt, Olivia Martius, Rolf Weingartner, Markus Mosimann, Veronika Röthlisberger, and Andreas Paul Zischg. 2019. "Characterizing precipitation events leading to surface water flood damage over large regions of complex terrain." *Environmental Research Letters* 14 (6): 064010.

Centraal Bureau voor de Statistiek. 2020. *Forecast: Population growth unabated in the next 50 Years.* https://www.cbs.nl/en-gb/news/2020/51/forecast-population-growth-unabated-in-the-next-50-years.

Dottori, Francesco, Peter Salamon, Alessandra Bianchi, Lorenzo Alfieri, Feyera Aga Hirpa, and Luc Feyen. 2016. "Development and evaluation of a framework for Global Flood Hazard Mapping." *Advances in Water Resources* 94:87–102.

Fabris, Fabio, Aoife Doherty, Daniel Palmer, João Pedro de Magalhães, and Alex A Freitas. 2018. "A new approach for interpreting Random Forest models and its application to the biology of ageing." *Bioinformatics* 34 (14): 2449–2456.

Feng, Wenxian, Chenkai Ma, Guozhang Zhao, and Rui Zhang. 2020. "FSRF:An Improved Random Forest for Classification." In *2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications(AEECA),* 173–178.

Classification: Precision and Recall. https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall.

Grahn, Tonje. 2020. *Assessment of Residential Flood Damage Functions to Guide Policy Choices,* 1–8.

Grahn, Tonje, and Lars Nyberg. 2017. "Assessment of pluvial flood exposure and vulnerability of residential areas." *International Journal of Disaster Risk Reduction* 21:367–375.

Guidotti, Riccardo, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2018. "A Survey of Methods for Explaining Black Box Models." *ACM Comput. Surv.* (New York, NY, USA) 51 (5).

Hastie, Trevor, Jerome Friedman, and Robert Tisbshirani. 2017. *The elements of Statistical Learning: Data Mining, Inference, and prediction.* Springer.

Holzinger, Andreas, Peter Kieseberg, Edgar Weippl, and A. Min Tjoa. 2018. "Current Advances, Trends and Challenges of Machine Learning and Knowledge Extraction: From Machine Learning to Explainable AI." In *Machine Learning and Knowledge Extraction,* edited by Andreas Holzinger, Peter Kieseberg, A Min Tjoa, and Edgar Weippl, 1–8. Cham: Springer International Publishing.

Jordan, M. I., and T. M. Mitchell. 2015. "Machine learning: Trends, perspectives, and prospects." *Science* 349 (6245): 255–260.

Karmaker ("Santu"), Shubhra Kanti, Md. Mahadi Hassan, Micah J. Smith, Lei Xu, Chengxiang Zhai, and Kalyan Veeramachaneni. 2021. "AutoML to Date and Beyond: Challenges and Opportunities." *ACM Comput. Surv.* (New York, NY, USA) 54 (8).

Kreibich, Heidi, and Annegret H. Thieken. 2008. "Assessment of damage caused by high groundwater inundation." *Water Resources Research* 44 (9).

Lamers, Christiaan, Jan van Rijn, and Ton Beenen. 2020. "Data science-technieken voor regenwateroverlast in stedelijk gebied." *H20-online.*

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep learning." *Nature* 521 (7553): 436–444.

Maddox, Ivan. 2014. *Three common types of flood explained.* https://www.intermap.com/risks-of-hazard-blog/three-common-types-of-flood-explained.

Moel, H. de, and J. C. Aerts. 2010. "Effect of uncertainty in land use, damage models and inundation depth on flood damage estimates." *Natural Hazards* 58:407–425.

What is a return period? 2014. https://niwa.co.nz/natural-hazards/faq/what-is-a-return-period.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. "Scikit-learn: Machine learning in Python." *Journal of machine learning research* 12:2825–2830.

Quax, Rick, Omri Har-Shemesh, and Peter Sloot. 2017. "Quantifying synergistic information using intermediate stochastic variables." *Entropy* 19 (2): 85.

Ronk, Ian C, Sander van Splunter, and Arjan Knibbe. 2022. "Predicting Flooding from Local Features."

Rosenzweig, Bernice R., Lauren McPhillips, Heejun Chang, Chingwen Cheng, Claire Welty, Marissa Matsler, David Iwaniec, and Cliff I. Davidson. 2018. "Pluvial flood risk and opportunities for resilience." *WIREs Water* 5 (6).

Rözer, Viktor, Meike Müller, Philip Bubeck, Sarah Kienzler, Annegret Thieken, Ina Pech, Kai Schröter, Oliver Buchholz, and Heidi Kreibich. 2016. "Coping with pluvial floods by private households." *Water* 8 (7): 304.

Simons, Thijs. 2021. *Wateroverlast.* https://github.com/SimonsThijs/wateroverlast.

Skougaard Kaspersen, Per, Nanna Høegh Ravn, Karsten Arnbjerg-Nielsen, Henrik Madsen, and Martin Drews. 2017. "Comparison of the impacts of urban development and climate change on exposing European cities to pluvial flooding." *Hydrology and Earth System Sciences* 21 (8): 4131–4147.

Spekkers, M. H., M. Kok, F. H. Clemens, and J. A. ten Veldhuis. 2014. "Decision-tree analysis of factors influencing rainfall-related building structure and content damage." *Natural Hazards and Earth System Sciences* 14 (9): 2531–2547.

Sunyer, Maria Antonia, Ida Bülow Gregersen, Dan Rosbjerg, Henrik Madsen, Jakob Luchner, and Karsten Arnbjerg-Nielsen. 2014. "Comparison of different statistical downscaling methods to estimate changes in hourly extreme precipitation using RCM projections from ensembles." *International Journal of Climatology* 35 (9): 2528–2539.

The Danish Emergency Management Agency. 2013. *National Risk Profile (NRP),* 4–14.

Touw, Wouter G., Jumamurat R. Bayjanov, Lex Overmars, Lennart Backus, Jos Boekhorst, Michiel Wels, and Sacha A. F. T. van Hijum. 2012. "Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle?" *Briefings in Bioinformatics* 14 (3): 315–326.

Turing, Alan M. 2009. "Computing Machinery and Intelligence." In *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer,* edited by Robert Epstein, Gary Roberts, and Grace Beber, 23–65. Dordrecht: Springer Netherlands.

Vaughan, Joel, Agus Sudjianto, Erind Brahimi, Jie Chen, and Vijayan N. Nair. 2018. *Explainable Neural Networks based on Additive Index Models.*

Wheater, Howard, and Edward Evans. 2009. "Land use, water management and future Flood Risk." *Land Use Policy* 26.

Appendices

A Case Study: Adding the Feature ‘Construction Year’

In this study, use has been made of two kind of local features to predict if a residence sustains rainwater damage; the rainfall and height measures. As defined in Table 2, it is also possible to add construction year as a feature to a model. In this case study the effect of adding this feature is investigated. This is done for a model that considers object-level water damage notifications in the P2000 network.

In the process of engineering a data set, as described in Section 4.1, it is possible to also add the construction year of a residence that experienced water damage. This information is retrieved using the National Building Register of the Netherlands. After engineering a data set, 10-fold cross-validation is applied on this data set to train and test random forest classifiers. K-fold cross-validation is widely used and reduces overfitting (Hastie, Friedman, and Tisbshirani 2017).

Results

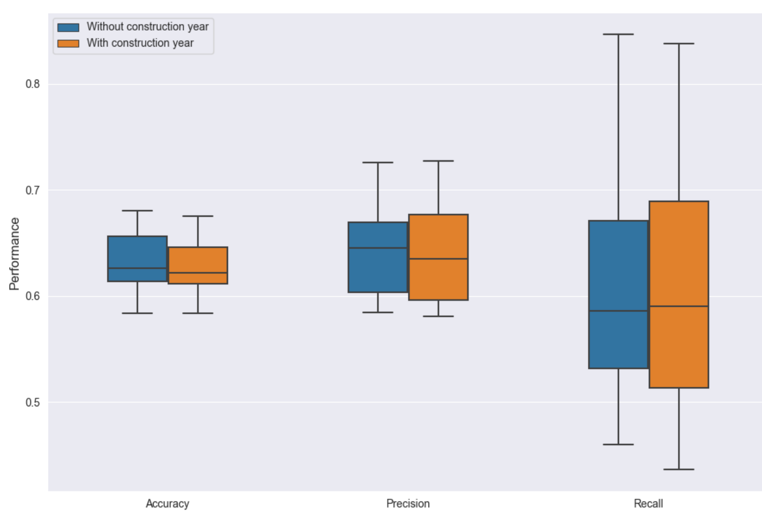

Figure 10: A box-plot with the distribution of evaluated performances of two created models, one with construction year as a feature and one not. 10-fold-cross validation is used as a test-method.

The results (Table 8 and Figure 10) show that for the defined task, the averages of accuracy, precision and recall are lower for random forest classifiers that also use construction year as a feature. Therefore, it is unlikely that a model that uses construction year as a predictive feature for rain damage performs better than a model that does not. At first sight, this seems contradicting with research performed by Spekkers et al. (2014). In his study, it is concluded that a buildings age is associated with the number of claims relating to rainwater damage (Spekkers et al. 2014). This difference in results can be caused by the different target variable. Predicting the number of

Random forest classifier	Average accuracy	Average precision	Average recall
without construction year as a feature	0.6325	0.6451	0.6139
with construction year as a feature (w.r.t. without)	- 0.0063	- 0.0038	- 0.0090

Table 8: The performance of two models that use a random forest classifier and 10-fold cross-validation. One includes construction year of a residence as a feature and one not.

insurance claims is a different task than predicting if someone notifies emergency services. Both of these are used as a proxy-variable for rainwater damage in general. This might suggest that rain damage to a residence depends more on the construction year of the residence than that rain nuisance in a residence is dependent on the construction year of the residence.