Informatica en Economie

Using machine learning techniques

to predict rainwater damage instances

Pi Molling

Student ID:
2847116

Supervisor:
J.N. van Rijn
External Supervisor:
T. Beenen (RIONED)

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl 14/08/2023

Abstract

In this thesis, we have built models to predict rainwater damage instances at three different spatial resolutions, respectively object level, subdistrict level and district level. Messages from the P2000 network, a communication network of emergency services in The Netherlands, are used to obtain positive rainwater damage instances. Hereby, the proximity sampling method is used to sample negative rainwater damage instances. This sampling method samples nearby residences as negative instances from the positive rainwater damage instances, where no rainwater damage was reported. Hereby, we feed the model a complex problem to learn because the positive and negative instances consequently will have the same amount of rainfall. For the problem at each resolution, we used random forest classifiers to make predictions for unseen instances. Based on the results, the models at the object level have the highest performance of *57.60%* accuracy followed by *55.66%* on the subdistrict level and *55.14%* on the district level. We notice a decreasing trend in the model's accuracy when a gradual loss of information occurs in the water damage data. Based on the results, we can conclude that our models at the object level achieve the highest performance.

Contents

1 Introduction

Flood is one of the common natural disasters that cause damage to residences in urban areas [1]. It can be caused by, e.g., heavy rainfall or overflowing of rivers [2]. In this study, we will build machine models that aim to predict rainwater damage instances, caused by heavy rainfall, in the Netherlands at different various resolutions.

1.1 Background

Urban areas in the Netherlands often have a high density of the country's population. They are the central hubs for economics and societal sectors [3]. Urban areas are prone to pluvial floods due to poor drainage systems [4]. Pluvial flooding can occur due to a combination of high intensity of rainfall and excessive burden on the drainage system's capacity. The excessive amount of water remains on the surface and leads to flooding of streets and buildings.
Floods in urban areas can be divided into four different types [5]:

1. *Pluvial Floods*: This is caused by high intensity of rainfall in a short period which exceeds the capacity of drainage systems;

2. *Fluvial Floods*: Fluvial floods are caused by the overflowing of rivers and streams due to heavy rainfall;

3. *Ground Water Floods*: Due to the rising groundwater level, the water exceeds the capacity of the soil to absorb or transport water. This results in the emergence of excessive water on the surface;

4. *Coastal Floods*: These floods happen in coastal zones where the natural flood barriers (dunes) are weak or not present. It is usually caused by storm surges.

The financial damages that are caused by flooding can be various, depending on the scale of the flood and the area it affects. In 2008, the estimated financial building damages and contents damages caused by floods in Lohmar are respectively €269,496 and €355,953 [3]. Eventhough pluvial flooding is common in urban areas, it gains less attention than fluvial flooding as its impacts are at a small scale [6] and do occur not often. Stukstette et al. [7] have proposed that projects regarding rainwater damages must be initialised within a policy agenda. For instance, residents in Osnabrück received an early warning of possible pluvial flooding and emergency measures could be taken to reduce damages to the properties and buildings [3]. It is advisable to consider how pluvial flooding can be prevented in certain vulnerable urban areas in the Netherlands. So, several solutions can be considered and implemented on time.

An important step that can be taken to prevent pluvial flooding is by developing predictive machine learning models that predict rainwater damage instances [8, 9, 10, 11]. Historical data regarding rainwater damages must be obtained in order to train and test the models [8, 9, 10, 11]. Twitter messages, P2000 notifications and insurance claims (Lamers, van Rijn and Beenen [9], Bavelaar [10], de Mast [8]) have been used to develop various models [8]. The P2000 network contains notifications about emergencies in the Netherlands. The previously developed models use rainfall and height layers as features [9, 10, 8]. In this study, we will develop several models using data from the P2000 network, the Current Dutch Elevation (Actueel Hoogtebestand Nederland, *AHN*) map, the Royal Netherlands Meteorological Institute (*KNMI*) and National Building Register (Basisregistratie Adressen en Gebouwen, *BAG*).

1.2 Water Management In The Netherlands

As a large part of the Netherlands lies below sea level, it is prone to flood disasters. The flood of 1953 was the greatest flood event that ever occurred. Severe storms and poor condition of dykes are the main causes [12]. This flood disaster illustrated that flood preventive measures must be taken into consideration. Hereby, The Dutch waterboards and Rijkswaterstaat have the responsibility to come up with strategies to mitigate flood events in The Netherlands, respectively fluvial flooding and coastal flooding [13].
The foundation RIONED supports knowledge development and dissemination for water management in urban areas [14]. According to RIONED, rainwater damage is one of the common problems in urban areas [14]. The foundation notices that a structural plan must be initialised to prevent rainwater damage in a long term in urban areas.

1.3 Research Question

The spatial resolution of rainwater damage instances can differ. Instances obtained from the P2000 network are at the object level, whilst the insurance claims studied by de Mast [8] are at the district level. In this thesis, we will approach the problem of predicting rainwater damages in the same spatial resolutions, as presented in the study of de Mast [8], with the more realistic sampling method that is provided by Simons [11], from hereon dubbed as proximity sampling. This sampling method provides a realistic overview of negative samples, as it relies on the nearby residences with consequently rainfall amount as the positive rainwater damage instances. Hence this method feeds a complex problem for the models to learn. The research question for this thesis is as follows:

RQ: What is the effect of lowering the resolution of rainwater damage instances on the model's performance? [8]

The assumption towards the *RQ* is that lowering resolutions will result in a lower model's performance due to information loss.

1.4 Contributions

In the previous research, de Mast [8] has introduced experiments to measure the performance of the models that predict rainwater damage instances at three spatial resolutions using

the dependent sampling method provided by Bavelaar et al [10]. The results of the study carried out by de Mast [8] show that resolutions of water damage have negligible impact on the models' performance.
This research will apply the proximity sampling method introduced by Simons [11] instead to the same three spatial resolutions as presented in the research of de Mast [8] to evaluate the models' performance. The reason for this is that we suspect that this sampling method presents the machine learning model with a more challenging problem, and that as such the loss in performance across resolution can be measured. Subsequently, we will use the values of the height map and past 3 hours of rainfall as predictive features for our models.

2 Related Work

In this section, we will provide an overview of academics researches regarding predicting flood events and rainwater damage instances.

2.1 Fluvial Flood Forecasting

According to Wilby et al. [15], climate change may have an impact on fluvial flooding. Hydraulic models are commonly used to simulate flood water levels and have high accuracy in predicting flood wave propagation [16]. However, these models require significant computational time and are not suitable for the prediction of real-time flood events [16]. On the other hand, a deep convolutional neural network (DCNN) is suitable for rapidly predicting fluvial flooding [17]. The results of the study of Kabir et al. [17] illustrate that the DCNN model achieves high accuracy results and can be retrained when changes appear in the topography.

2.2 Predicting Pluvial Flooding

Qian Ke et al. [18] have used rainfall intensities to classify flood events in their machine learning approach for a case study in Shenzhen, China. This study suggests two thresholds for two different temporal resolutions, respectively a threshold of 20mm for 30 minutes of rainfall and 80mm for 3 hours of rainfall. When the threshold is exceeded, the event is regarded as a pluvial flood [18].
Zahura et al. [19] build a machine learning surrogate model using a random forest algorithm to imitate the physics-based model for forecasting pluvial flooding in urban areas of Norfolk, USA. This approach requires less computational time than the conventional physics-based model. The surrogate model required 56 minutes for training and 4.2 seconds to predict each test event, whilst the physics-based model requires 4.5 to 6 hours per execution [19].

2.3 Predicting Rainwater Damage Instances

Bavelaar et al. [10] have combined rainfall data, and height map data with P2000 and Twitter messages to build predictive models. Hereby, Bavelaar et al. [10] concluded that models using P2000 messages are more reliable as P2000 messages contain less noise than Twitter messages. Thereby, three different sampling methods have been proposed to sample negative instances, as messages from the P2000 network contain only positive instances of rainwater damage. Additionally, Simons [11] introduced another sampling method in his study: proximity sampling. The definitions of the four sampling methods are as follows:

1. *Random*: This method samples negative instances using random coordinates within The Netherlands. The amount of rainfall is dependent on the sampled coordinates [10]. Hereby, the sampled negative instances could be randomly selected points on streets, houses, farmland or buildings.

2. *Equal location/Dependent*: The negative and positive instance has the same location. The negative instance is taken from a different timestamp and its rainfall amount is randomly generated between the threshold and the highest amount of rainfall in the train data [10].

3. *Address*: The negative instances are sampled based on random addresses in The Netherlands obtained from National Building Register [10].

4. *Proximity*: This sampling method is based on addresses in the Netherlands. Hereby, the sampled negative instances are nearby residences from the positive instances and will have consequently the same amount of rainfall as the positive rainwater damage instances [11].

De Mast [8] has built models at three different spatial resolutions, respectively object level, subdistrict level and district level. Hereby, the dependent sampling method of Bavelaar et al. [10] has been used. Notably, the differences between the accuracy performances across the models at different spatial resolutions are less than 1% [8]. By using dependent sampling as an evaluation method, de Mast concluded that the spatial resolution of water damage data has negligible impact on the model's performance [8].

Using the proximity sampling method in this study, the model aims to answer the following question: "*Why do some residences experience rainwater damage at a given timestamp and other nearby residences not?*"

This research contributes to the previous works that have been conducted by Bavelaar et al., Simons and de Mast [8, 10, 11].

3 Data

This section aims to provide an overview of the data gathering methods that are used in this study for the purpose of creating train data for machine learning models. All obtained data are open source. All data sources are presented in Table 1. First, we will describe the rainwater damage instances, the target variable, followed by rainfall and height features.

3.1 Rainwater Damage Instances

In this study, only notifications from the P2000 network will be used to label rainwater damage instances. P2000 is a part of the C2000 network, a communication network of emergency services in The Netherlands. The usage of API is required to obtain the notifications. Hereby, we used a script provided by Simons [11] and later used by de Mast [8] to retrieve notifications with messages that are related to water nuisance, which likely can be instances of water damage. The obtained emergency notifications are in JSON format, an example is shown in Figure 1. The emergency notifications provide information such as timestamp, address information, postal code, latitude, longitude, and province. A total of 5848 notifications were retrieved to serve as positive instances, meaning we have complete address information. A key observation to point out is that all emergency instances are at the object level, including instances at subdistrict and district could yield in a more extensive data points. Also, there is uncertainty whether each of these notifications is directly related to the water damage caused by rainfall. This might result in noise in both train and evaluation data. A measure has been taken to avoid this problem. The data will be filtered to contain instances where the amount of the past 3 hours of rainfall exceeds the threshold of 50mm proposed by Simons [11]. The choice for this threshold is to ensure that the damage is caused by rainfall and not, for example, a broken pipe [11]. This threshold is arbitrarily chosen and could be further optimised [11]. Lastly, the filtered instances are labelled as positive instances affected by rainwater damage.


```
{
        "prio": 2,
        "message":"P2 7234 Zonneplein 63 Bergen op Zoom Wateroverlast",
        "service": "Brandweer",
        "date": "2016-01-04 09:00:53",
        "region": "Midden- en West-Brabant",
        "region_url": "midden-en-west-brabant",
        "place": "Bergen op Zoom",
        "place_url": "bergen-op-zoom",
        "address": "Zonneplein 63, Bergen op Zoom",
        "zipcode": "4624",
        "province": "Noord-Brabant",
        "latitude": 51.4857216,
        "longitude": 4.3098238,
}
```


Figure 1: Notification from P2000 network.

#	Data Source	Temporal Resolution	Spatial Resolution	Period	Publicly available
1	P2000 network	Exact Timestamp	Object Level	2016-2022	Yes
2	Current Elevation Map of the Netherlands (AHN3)	1 scan	0.5m $\times$ 0.5m pixels	Snapshot	Yes
3	Precipitation dataset from The Royal Meteorological Institute of The Netherlands	Every 5 minutes	1km	2016-2022	Yes
4	Database from Statistics Netherlands	1 scan	-	2021	Yes
5	National Building Register	1 scan	Object Level	Snapshot	Yes

Table 1: Overview of data sources used in this study modified from de Mast [8]
.

3.2 Rainfall Feature

One of the features in our train data is the *'3 hours rainfall'* feature. This feature implies the amount of rainfall 3 hours preceding the timestamp of incoming P2000 notifications. It has been used by Bavelaar et al. [10] and de Mast [8] to build the models. The rainfall data is provided by the Royal Meteorological Institute of The Netherlands (*KNMI*). This data is measured every 5 minutes at 1 kilometre grid by using radar technology [20]. Hence, residences situated in the same 1 kilometre grid have the same amount of rainfall. The data is stored in Hierarchical Data Format, version 5 (HDF5) which supports large and complex data [21]. It can be read either by using the h5py library or HDF view tools.

3.3 Height Features

In this study, we have added height features to our train data. These features are obtained using the Current Elevation Map of The Netherlands version 3 (*AHN3*). In total, there are 4 versions of the elevation map of the Netherlands: “AHN1”, “AHN2”, “AHN3” and “AHN4” [22]. The height maps can either be downloaded or accessed by API calls.
We have used the method to retrieve the height maps of “AHN3” from Simons [11]. The retrieved height map has the size of 10m $\times$ 10m. De Mast [8] has concluded that using height maps larger than 10m $\times$ 10m could result in a worse model's performance. The obtained height maps are in TIFF images format which can be read using the GDAL library. Each pixel

on the map covers an area of 0.5m $\times$ 0.5m and has one height value. This results in an array of 400 height values [8, 11]. All these values are added to the train data as features. The method to retrieve the height map is provided by Simons [11]. The choice for this method is to reduce the storage of large files on the local machine because at least 1TB of storage is needed for the entire AHN3 data [23].

Variables	Value type	Source
Target		
Rainwater damage instances	Binary	1
Features		
3 hours rainfall	Continuous	2
Height values (1 . . . 400)	Continuous	3

Table 2: Overview of features used in this study to predict rainwater damage instances target. Modified from de Mast [8]

.

4 Methods

In this study, we created three models to answer the research question, “*What is the effect of lowering the resolution of rainwater damage instances on model performance?*” [8]. Each model has a different spatial resolution, respectively object level, subdistrict level and district level. In this study, we will use the proximity sampling method to sample negative instances of positive rainwater damage instances. This section provides an overview of the proximity sampling method and our machine learning pipeline including the steps for obtaining instances of the negative class and building machine learning models.

4.1 Proximity Sampling

The filtered P2000 notifications, described in Section 3.1, provide the positive instances in the train data. Hereby, the negative instances must be synthetically sampled. This study has opted for the proximity sampling method developed by Simons [11]. This sampling method is suitable for this study as it ensures that the negative instances are nearby residences rather than street instances or random residences within the Netherlands, shown in Figure 2. Notice that the negative rainwater damage instances will consequently have the same amount of rainfall as the positive instances. This sampling method gives us a realistic overview of the question "*Why do some residences experience rainwater damage at a given amount of rainfall and other nearby residences not?*" and serves a complex problem to the models. To perform this sampling method, data from the National Building Register is required.

Figure 2: Example of 1 positive rainwater damage instance, coming from the P2000 network, with its sampled negative instance by applying proximity sampling method. Because they are close to each other they will have similar amounts of rainfall. This forces the model to differentiate on height features.

4.2 Resolutions

This section will provide an overview of how train data has been created at three different spatial resolutions: object level, subdistrict level and district level. This is the first part in our machine learning pipeline, illustrated in Figure 3. Note that the process of engineering the train data of each spatial resolution is different. The final train data of all spatial resolutions are balanced, meaning that the amount of positive instances is equal to the amount of negative instances.

4.2.1 Object Level

The retrieved P2000 notifications described in Section 3.1 represent object level instances. These notifications contain complete information about the postal code and house number of water damage instances. The following steps will be taken to engineer the train data:

1. *Add past 3 hours rainfall*:
   The accumulated rainfall of the past three hours preceding the timestamp of incoming P2000 water damage notifications will be calculated and added to each record, using the precipitation data from The Royal Meteorological Institute of The Netherlands, described in Section 3.2 [8, 11].

2. *Filter rainwater damage instances*:
   The data will further be filtered using the threshold of 50mm suggested by Simons [11]. The result of the filtered data is classified as positive rainwater damage instances.

3. *Add height layers*:
   The height layers for each rainwater damage instance will be obtained using The Current Elevation Map of The Netherlands. We are interested in an area of 10m by 10m around each object as described in Section 3.3. This results in an array of 400 values [8]. These values are added as features [8, 11]. However, in future work, a different area could be explored.

4. *Add negative instances*:
   The proximity sampling method designed by Simons [11] will be used to sample negative instances of rainwater damage, based on the assumption when there is no message in the P2000 dataset at a given timestamp, there was no rainwater damage. Note that the sampling approach differs for each resolution, illustrated in Table 4. The approach for each resolution is as follows:

   (a) *Object Level*:
   For each rainwater damage instance at the object level, we will randomly sample an address in the same subdistrict. This address is used as a negative instance. This is done because it is possible that nearby residences in the same subdistrict

may not experience rainwater damage.

(b) *SubDistrict Level*:
At the subdistrict level, we will assume that we do not know the exact house number. As such, all residences in a certain subdistrict are classified as positive rainwater damage instances. Hereby, one random address will be used as a rainwater damage instance from the subdistrict provided by the P2000 notification. Hereby, the sampled negative instance of rainwater damage at the subdistrict level should be located in another subdistrict (within the same district) from the positive instance.

(c) *District Level*:
Notifications at the district level imply that a certain residence in a district experienced rainwater damage, as we will assume that we do not know the complete postal code and house number information. At the district level, there is a higher level of uncertainty about which residences experienced rainwater damage. Likewise, at the subdistrict level, one address will be used as a rainwater damage instance from a district provided by P2000 notification. Hereby, the sampled negative instance at the district level of rainwater damage should be located in another district from the positive instance.

4.2.2 SubDistrict Level

At the subdistrict level, we assume that the incoming P2000 notifications do not contain the house numbers. Subdistrict instances will be transformed into object instances. A random existing residence will be chosen from the data provided by the National Building Register for each subdistrict instance. Hereby, a random address is used as a water damage instance for each subdistrict instance. Then, the steps described in Section 4.2.1 are proceeded to engineer the final train data.

4.2.3 District Level

At the district level, we assume that the notifications from the P2000 network contain only 4 digit postal code. Hereby, the data provided by Statistics Netherlands is used to choose a random subdistrict from the 4 digits postal code. Thereafter, a random house number will be chosen by using data from the National Building Register. Hereby, a random address is used as water damage instance for each district instance. Then, the steps described in Section 4.2.1 are taken to create the final train data.

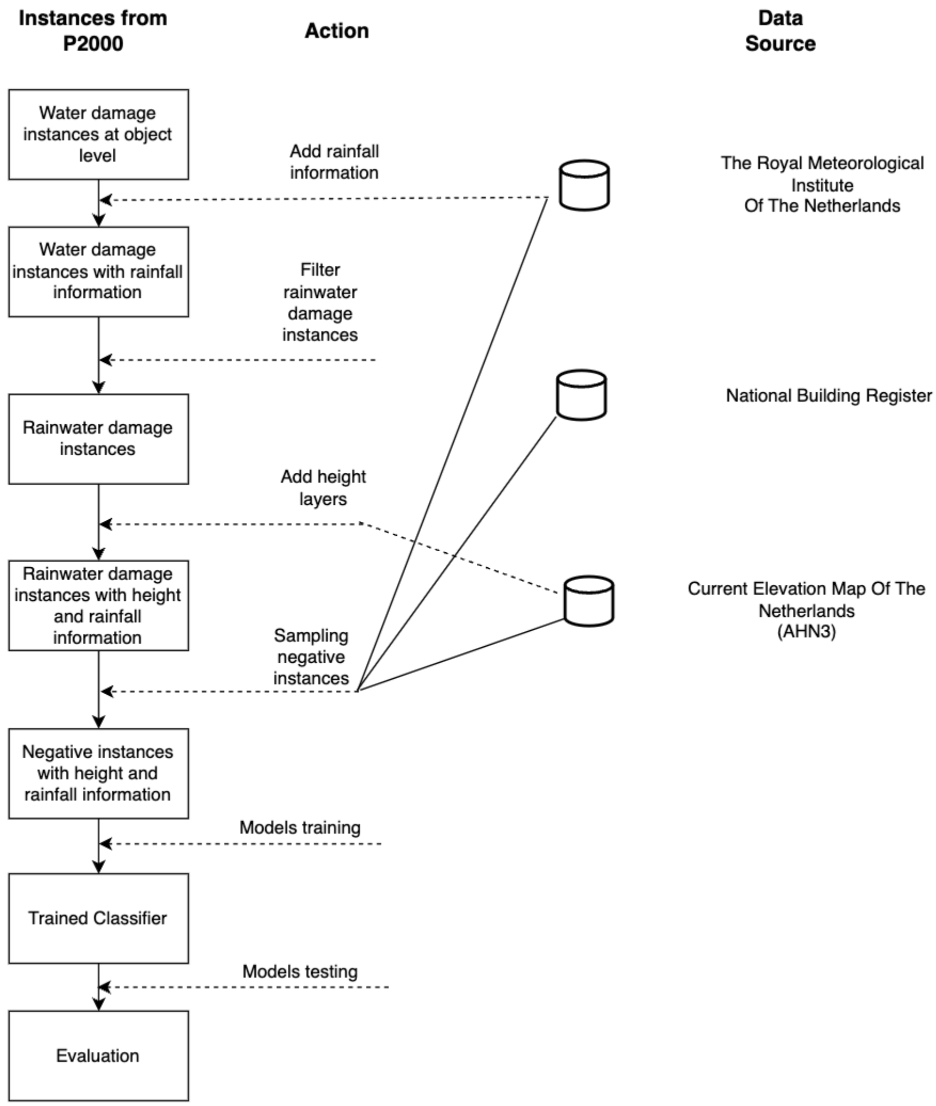


Figure 3: Machine learning pipeline modified from de Mast [8].

4.3 Machine Learning Models

After generating train data described in Section 4.2, a machine learning algorithm will be chosen to train to create models. Various machine learning algorithms can be used to create models. In the research setting, random forest is a popular choice among other machine learning algorithms [24]. It is efficient, requires less hyperparameters tuning than other algorithms and can handle large amounts of data and has high accuracy for several types of data [25]. Therefore, random forest algorithms will be used in this study to create classification models. Table 3 presents the set of hyperparameters used for random forest algorithms.

4.4 Models Training and Testing

After engineering our data with features and negative instances, and selecting a machine learning algorithm, the subsequent steps in our machine learning pipeline involve training and testing the models, as presented in Figure 3. In the train data, each positive rainwater damage instance is grouped with its corresponding negative instance, effectively making them end up together either in the train set or test set. This results in approximately 1000 groups in our data. To mitigate bias in the models' performances, a 10-fold group cross-validation technique is used [26]. Hereby, 90% of the data will be allocated for the training stage, while 10% will be used during the testing stage. As such, each random forest classifier will be trained and tested using 10 different subsamples of the data. Notice that each group will be assigned for either training or testing purposes in each iteration [27]. Figure 4 illustrates the training and testing group of data in each iteration of 5-fold group cross-validation.

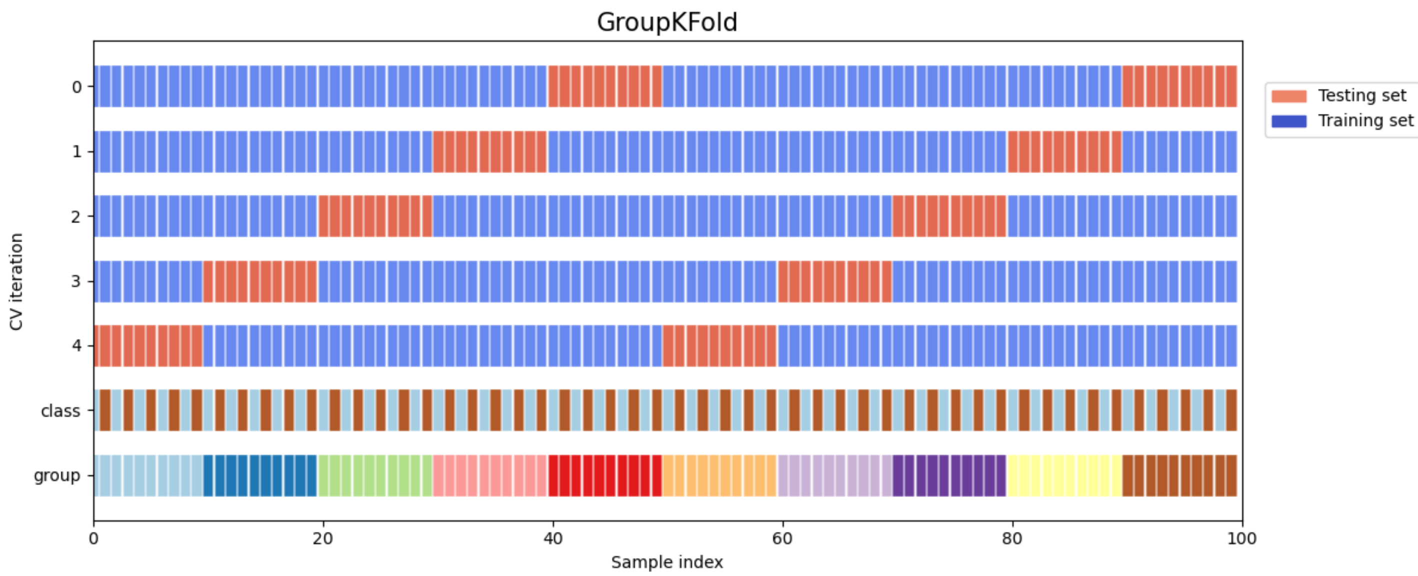


Figure 4: Overview of 5-fold group cross-validation modified from Pedregosa et al. [28]. Note that we use 10-fold group cross-validation in this study.

Hyperparameter	Value
Numbers of trees in the forest	1000
Maximum depth of each decision tree	5
Minimum number of samples to split an internal node	2
Minimum number of samples to be present in a leaf	1

Table 3: Overview of hyperparameter used for random forest algorithms.

4.5 Evaluation Metrics

The evaluation process occurs at the testing stage of the 10-fold group cross-validation, outlined in Section 4.4 and presented in Figure 3. In this study, the following metrics are used to measure the performances of the random forest classifiers:

1. *Accuracy*:
The ratio of correctly predicted instances to the total number of evaluated classes across instances [29], given by the following formula:

$$\frac{\textit{Number Of Correct Predictions}}{\textit{Total Number Of Predictions}}$$

2. *Precision*:
The ratio of correctly predicted positive classes to the number of instances in the predicted positive predicted class [29], given by the following formula:

$$\frac{\textit{True Positives}}{\textit{True Positives} + \textit{False Positives}}$$

3. *Recall*:
The ratio of correctly predicted instances to the number of actual positive instances [29], given by the following formula:

$$\frac{\textit{True Positives}}{\textit{True Positives} + \textit{False Negatives}}$$

5 Results

In this section, the results of the conducted experiments will be presented. Section 5.1 will provide visualizations of data at the object, subdistrict and district level, followed by the results of the experiments conducted under the normal sample range setting and district sample range setting. Lastly, followed by the results of statistical tests conducted on the results of the district sample range setting.

5.1 Data Visualizations

Figure 5 shows an overview of positive rainwater damage and negative instances on a map after engineering train data at the object level. Using the threshold of 50mm rainfall in the past three hours preceding the incoming P2000 messages, we notice that approximately 83% of water damage instances are filtered out. The remaining 17% are classified as positive rainwater damage instances.

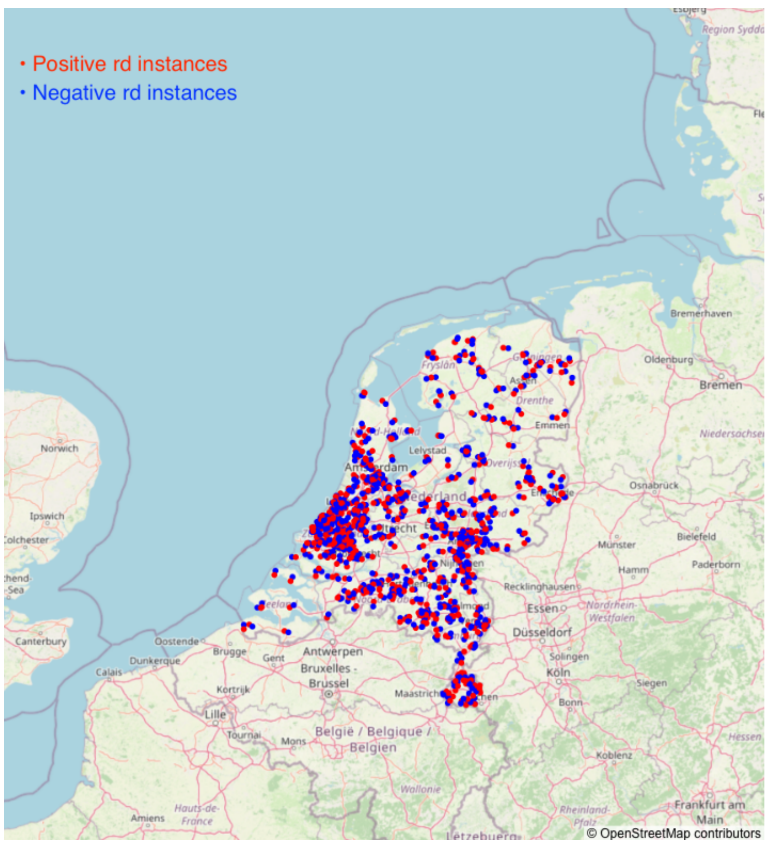
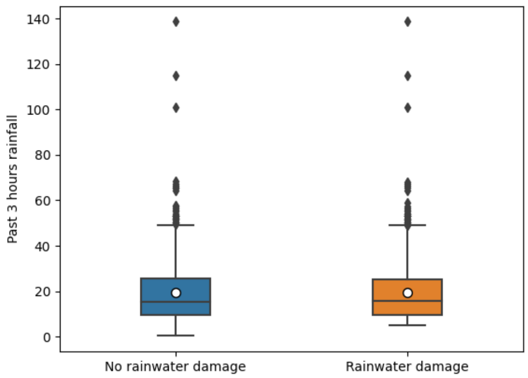
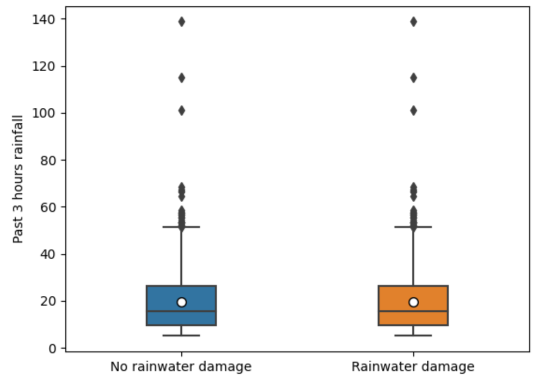


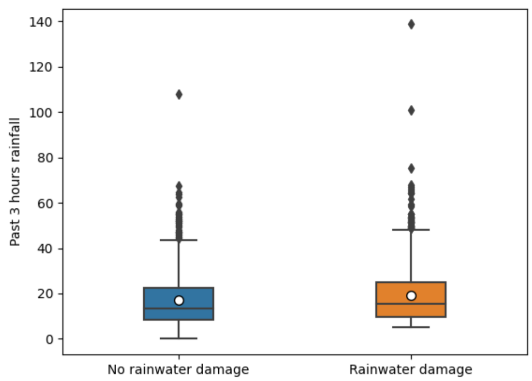
Figure 5: Overview of positive and negative rainwater damage instances at the object level using the proximity sampling method.



(a) Object level



(b) Subdistrict level



(c) District level

Figure 6: Bar plots showing the distribution of past 3 hours rainfall for positive and negative at various spatial resolutions.

Figures 6a and 6b emphasise that positive and negative rainwater damage instances at the object and subdistrict level approximately have the same amount of rainfall in the past three hours preceding experiencing rainwater damage. Notably, there are significant differences in the amount of rainfall between positive and negative rainwater damage instances at the district level, presented in Figure 6c.

5.2 Results Of Predictions In The Normal Sample Range Setting

The first experiment has been conducted under the normal sample range setting, outlined in Section 4.2 and presented in Table 4. Table 5 and Figure 7 present the results of this experiment setting.

Spatial Resolution	Positive instances	Negative instances
Object Level	Object level	Same SubDistrict
SubDistrict Level	SubDistrict	Other SubDistrict
District Level	District	Other District

Table 4: Overview of positive and negative locations of rainwater damage at various spatial resolutions in the normal sample range setting.

Spatial Resolution	Average accuracy	Average precision	Average recall
Object Level	0.5695	0.6001	0.4209
SubDistrict Level	0.5307	0.5317	0.5035
District Level	0.5514	0.5450	0.6263

Table 5: Cross-validated prediction results on various test sets with rainfall and height features in the normal sample range setting at various spatial resolutions of rainwater damage instances. Hereby, the evaluation metrics discussed in Section 4.5 are used.

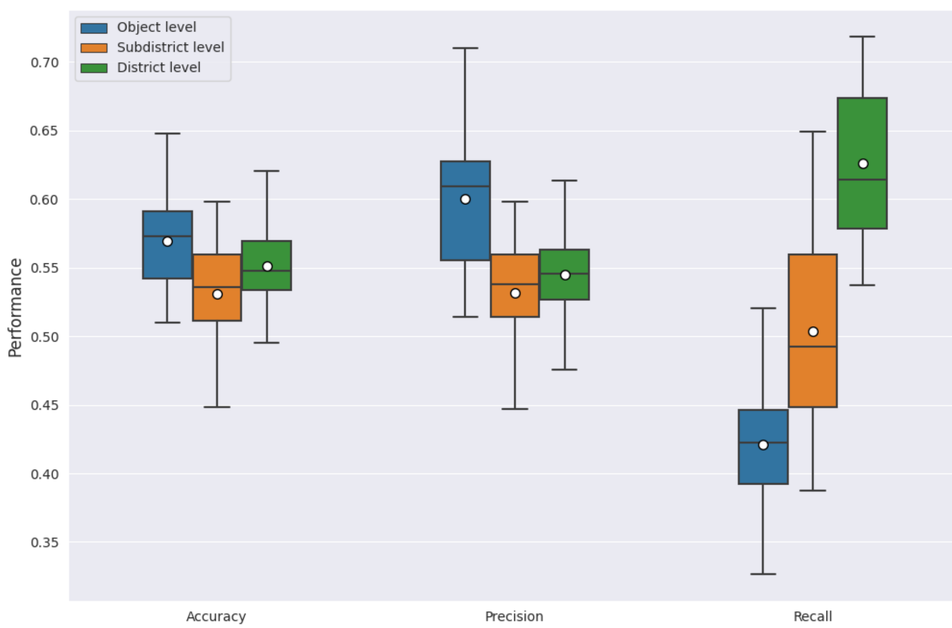


Figure 7: A box plot of the cross-validated prediction results of classifiers at various spatial resolutions of rainwater damage instances. Hereby, height values and the past three hours rainfall are the predictive features.

Table 5 and Figure 7 present the following evaluation metrics for the random forest classifiers: accuracy, precision and recall as described in Section 4.5. Hereby, the classifiers

at the object level have the highest accuracy and precision score compared to classifiers at the (sub)district level. However, classifiers at the subdistrict level in this setting have lower accuracy scores than classifiers at the district level. Our assumption that information loss will lead to a decrease in the performance of the classifiers seems therefore incorrect. However, there might be a logical reason for this.
We assume that there is a high difference in distance between positive and negative rainwater damage instances at the district level, as the negative instances are sampled in another district other than where the positive instances are located.[1] This leads to higher differences in past 3 hours rainfall between the positive rainwater damage instances and negative rainwater damage instances at the district level, as illustrated in Figure 6c. Hence, high differences in rainfall could enhance the model's ability to predict rainwater damage instances, as it does no longer have to rely solely on height features. To further investigate this, we have conducted the second experiment setting where we increase the distance between positive and negative rainwater damage instances at all spatial resolutions: the district sample range setting.

5.3 Results Of Prediction In The District Sample Range Setting

In this section, we will discuss the results of the second experiment: the district sample range setting. This experiment has been set up based on the assumption in Section 5.2 that high differences in past 3 hours rainfall at the district level could enhance the model's performance to predict rainwater damage instances, as the negative instances are sampled in another district other than where the positive instances are located.[1] Hence the models at the district level no longer need to rely solely on height features. Therefore, classifiers at the district level have higher accuracy scores than the classifiers at the subdistrict level. In this experiment setting, we aim to equalize the sample range of negative rainwater damage instances at each spatial resolution. So negative rainwater damage instances will be sampled in another district from positive rainwater damage instances across all spatial resolutions, shown in Table 6. Table 7 and Figure 8 present the results of this experiment setting.

Spatial Resolution	Positive instances	Negative instances
Object Level	Object level	Other District
SubDistrict Level	SubDistrict	Other District
District Level	District	Other District

Table 6: Overview of positive and negative locations of rainwater damage at various spatial resolutions in the district sample range setting.

[1]Note: the rainfall data is measured at 1 kilometre grid

Spatial Resolution	Average accuracy	Average precision	Average recall
Object Level	0.5760	0.5870	0.5108
SubDistrict Level	0.5566	0.5519	0.6119
District Level	0.5514	0.5449	0.6263

Table 7: Cross-validated prediction results on various test sets with rainfall and height features in the district sample range setting at various spatial resolutions of rainwater damage instances. Hereby, the evaluation metrics discussed in Section 4.5 are used.

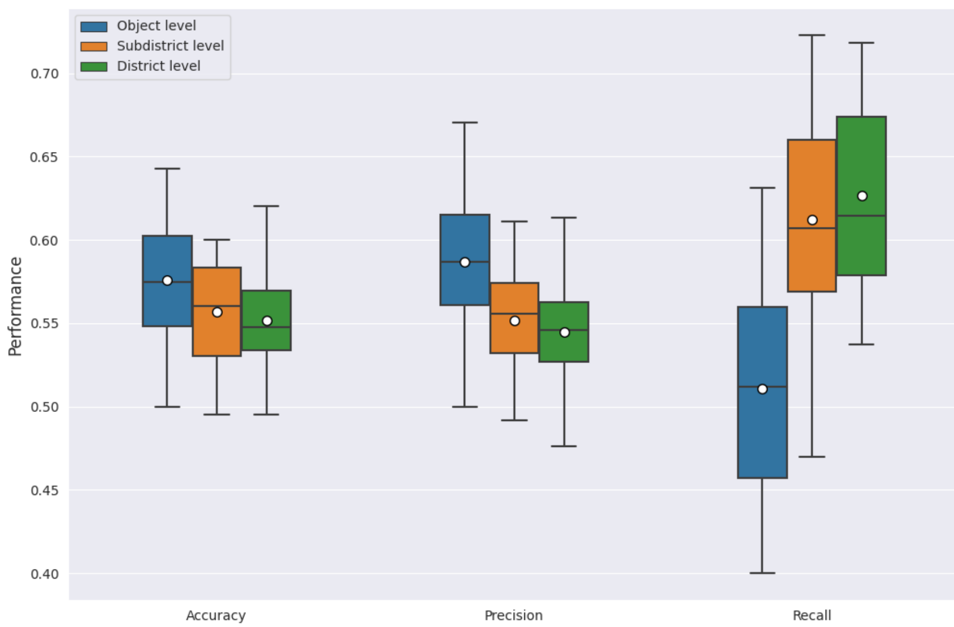


Figure 8: A box plot of the cross-validated prediction results of classifiers at various spatial resolutions of rainwater damage instances. Hereby, height values and the past 3 hours rainfall are the predictive features.

Table 7 and Figure 8 present the following evaluation metrics for the random forest classifiers: accuracy, precision and recall as described in Section 4.5. In this setting, the classifiers at the object level have the highest accuracy score followed by classifiers at the subdistrict and district level. Hereby, a decreasing trend of accuracy scores can be noticed when the classifiers suffer from information loss. Also, we noticed that classifiers at the object level have the highest precision scores but the lowest recall scores compared to classifiers at the (sub)district level in both experiment settings.

The results of average accuracy at three different spatial resolutions, presented in Table 7, are used to perform an ANOVA test. This test determines whether there is a statistically significant difference between the average accuracy of the classifiers [30].

H_0 = The average accuracy of the three classifiers is equal.
H_a = At least one average accuracy is different.

Hereby, a P-value of 0.05 is used to determine statistical significance. The p-value of the test result is 0.009. As the p-value of the test result is below 0.05, we can reject the null hypothesis.

To determine which average accuracy results are different, we performed the Nemenyi test. The results of the Nemenyi are presented in Table 8. Using an α-value of 0.05, the only statistical difference between the average accuracy is observed between classifiers at the object level and district level.

	Object Level	SubDistrict Level	District Level
Object Level	1.000000	0.162824	0.031943
SubDistrict Level	0.162824	1.000000	0.772157
District Level	0.031943	0.772157	1.000000

Table 8: Nemenyi test results of the average accuracy of random forest classifiers at three different spatial resolutions in the district sample range setting.

6 Discussion

We have examined the performances of random forest classifiers at three different spatial resolutions by conducting two experiments. Both experiments include rainfall and height features for the random forest classifiers. In the first experiment, normal sample range setting, outlined in Section 5.2, the result shows that the classifiers at the object level have the best performance of *56.95%* average accuracy. However, we notice that the classifiers at the district level have higher accuracy than the classifiers at the subdistrict level. The negative instances at the district level lie in another 1 kilometre grid where the amount of rainfall could have higher differences from the positive instances, as explained in Section 5.2. Hereby, high differences in the past 3 hours rainfall between positive and negative rainwater damage instances at the district level could enhance the models' performance to predict rainwater damage instances.
In the second experiment, the district sample range setting, the classifiers at the object level still perform the best performance of *57.60%* accuracy followed by *55.66%* at the subdistrict level and *55.14%* at the district level. Hereby, we notice that lowering the resolution results in a decreasing trend of the classifier's average accuracy performance. In both settings results, we have seen that the classifiers at the object level have higher average precision scores than other classifiers at (sub)district levels but have lower average recall scores than classifiers at (sub)district levels. This phenomenon is also called precision–recall trade-off

within the machine learning field [31]. After performing statistical tests of the results in the district sample range setting, we found that there is only a statistically significant difference in average accuracy between classifiers at the object level and classifiers at the district level.

6.1 Limitations

One of the main limitations of this study is that data verification of train and evaluation data is not conducted after generating them. This could lead to noise in both train and evaluation data.

Also, negative instances of rainwater damage in this study are sampled randomly. This could lead to false negative instances in the created train and evaluation data.

The height features in the train data are high dimensional, as the height values are retrieved from TIFF images. A large portion of high dimensional data is often not informative for classification [32]. This affects the classification performance of random forest classifier [32].

7 Conclusion and Further Research

In this study, we have built three classifiers to answer the research question: "*What is the effect of lowering the resolution of rainwater damage instances on the model's performance?*" [8]. The P2000 network is the only source of rainwater damage instances. We have used the proximity sampling method provided by Simons [11] to obtain residence instances for a realistic overview of negative instances at each spatial resolution, as discussed in Section 4.1. Random forest algorithm has been used to develop predictive models at three different spatial resolutions, respectively object level, subdistrict level and district level. Hereby, the rainfall and height features are used as predictive features. Based on the experiments conducted in the normal sample range setting and district sample range setting, we can conclude that lowering spatial resolution leads to a decrease in the accuracy performance of the classifiers. Furthermore, a precision-recall trade-off can be noticed in the performance of classifiers at the object level. Only statistical significance is found between classifiers at the object level and classifiers at the district level. Based on the results, we can conclude that our models achieve the highest performance when complete information is presented in the P2000 messages.

In this research, we have used the threshold provided by Simons [11]. However, the threshold used in the case study in Shenzhen [18] could also be applied in further study. Also, the performance differences between machine learning models and deep learning models could be further studied for this problem. Lastly, to reduce irrelevant features in the height features, stratified sampling for feature subspace selection could be used to extract relevant features for each subspace [33].

References

[1] R. Price and Z. Vojinovic, "Urban flood disaster management," *Urban water journal*, vol. 5, no. 3, pp. 259–276, 2008.

[2] H. H. Loc, E. Park, D. Chitwatkulsiri, J. Lim, S.-H. Yun, L. Maneechot, *et al.*, "Local rainfall or river overflow? Re-evaluating the cause of the great 2011 Thailand flood," *Journal of Hydrology*, vol. 589, pp. 1–3, 2020.

[3] V. Rözer, *Pluvial flood loss to private households*. PhD thesis, Universität Potsdam, 2018.

[4] Y. Chen, H. Zhou, H. Zhang, G. Du, and J. Zhou, "Urban flood risk warning under rapid urbanization," *Environmental research*, vol. 139, pp. 3–10, 2015.

[5] A. K. Jha, R. Bloch, and J. Lamond, *Cities and flooding: a guide to integrated urban flood risk management for the 21st century*. World Bank Publications, 2012.

[6] D. Houston, A. Werrity, D. Bassett, A. Geddes, A. Hoolachan, and M. McMillan, "Pluvial (rain-related) flooding in urban areas: the invisible hazard," 2011.

[7] J. Stukstette, "Regen in de stad," 2012.

[8] T. de Mast, "Lowering the resolution of damage data in a model that predicts rainwater damage," 2022.

[9] C. Lamers, J. van Rijn, and T. Beenen, "Data science-technieken voor regenwateroverlast in stedelijk gebied." https://www.h2owaternetwerk.nl/vakartikelen/data-science-technieken-voor-regenwateroverlast-in-stedelijk-gebied, 2020.

[10] C. Bavelaar, T. Simons, J. van Rijn, M. Baratchi, and T. Beenen, "Toepassing van machine learning-technieken op voorspelling van regenwateroverlast in stedelijk gebied," 2021.

[11] T. Simons, "Water overlast: A github repository regarding predicting rainwater damage." https://github.com/SimonsThijs/wateroverlast, 2021.

[12] "The flood of 1953." https://www.rijkswaterstaat.nl/en/water/water-safety/the-flood-of-1953.

[13] R. de Graaf, N. van de Giesen, and F. van de Ven, "Alternative water management options to reduce vulnerability for climate change in the Netherlands," *Natural hazards*, vol. 51, pp. 407–422, 2009.

[14] "Regenwateroverlast in bebouwd gebied: voorkom lokale problemen." https://www.riool.net/regenwateroverlast-in-bebouwd-gebied-voorkom-lokale-problemen, 2016.

[15] R. L. Wilby, K. J. Beven, and N. Reynard, "Climate change and fluvial flood risk in the UK: more of the same?," *Hydrological Processes: An International Journal*, vol. 22, no. 14, pp. 2511–2523, 2008.

[16] S. E. Brownell, D. S. Hekmat-Scafe, V. Singla, P. Chandler Seawell, J. F. Conklin Imam, S. L. Eddy, T. Stearns, and M. S. Cyert, "A high-enrollment course-based undergraduate research experience improves student conceptions of scientific thinking and ability to interpret data," *CBE—Life Sciences Education*, vol. 14, no. 2, pp. ar21, 2, 2015.

[17] S. Kabir, S. Patidar, X. Xia, Q. Liang, J. Neal, and G. Pender, "A deep convolutional neural network model for rapid prediction of fluvial flood inundation," *Journal of Hydrology*, vol. 590, pp. 1–3, 2020.

[18] Q. Ke, X. Tian, J. Bricker, Z. Tian, G. Guan, H. Cai, X. Huang, H. Yang, and J. Liu, "Urban pluvial flooding prediction by machine learning approaches–a case study of Shenzhen city, China," *Advances in Water Resources*, vol. 145, p. 103719, 2020.

[19] F. T. Zahura and J. L. Goodall, "Predicting combined tidal and pluvial flood inundation using a machine learning surrogate model," *Journal of Hydrology: Regional Studies*, vol. 41, p. 101087, 2022.

[20] "Precipitation - 1 hour precipitation accumulations from climatological gauge-adjusted radar dataset for the Netherlands (1 km) in KNMI HDF5 format - KNMI Data Platform," 2021.

[21] "Hierarchical Data Formats - What is HDF5?." https://www.neonscience.org/resources/learning-hub/tutorials/about-hdf5.

[22] Ahn.nl, "Kwaliteitsbeschrijving." https://www.ahn.nl/kwaliteitsbeschrijving.

[23] Ahn.nl, "Distributie." https://www.ahn.nl/distributie.

[24] A. B. Shaik and S. Srinivasan, "A brief survey on random forest ensembles in classification model," in *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2018, Volume 2*, pp. 253–260, Springer, 2019.

[25] Y. Qi, "Random forest for bioinformatics," in *Ensemble machine learning: Methods and applications*, pp. 307–323, Springer, 2012.

[26] S. Yadav and S. Shukla, "Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification," in *2016 IEEE 6th International conference on advanced computing (IACC)*, pp. 78–83, IEEE, 2016.

[27] Z. Mahmood and S. Khan, "On the use of k-fold cross-validation to choose cutoff values and assess the performance of predictive models in stepwise regression," *The International Journal of Biostatistics*, vol. 5, no. 1, pp. 2–7, 2009.

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[29] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International journal of data mining & knowledge management process*, vol. 5, no. 2, p. 1, 2015.

[30] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine Learning Research*, vol. 7, pp. 1–3, 2006.

[31] M. Buckland and F. Gey, "The relationship between recall and precision," *Journal of the American society for information science*, vol. 45, no. 1, pp. 12–19, 1994.

[32] B. Xu, J. Z. Huang, G. Williams, Q. Wang, and Y. Ye, "Classifying very high-dimensional data with random forests built from small subspaces," *International Journal of Data Warehousing and Mining (IJDWM)*, vol. 8, no. 2, pp. 44–63, 2012.

[33] Y. Ye, Q. Wu, J. Z. Huang, M. K. Ng, and X. Li, "Stratified sampling for feature subspace selection in random forests for high dimensional data," *Pattern Recognition*, vol. 46, no. 3, pp. 769–787, 2013.

A Case Study: Postal Code 5

Additionally, we have examined the model's performance at the postal code 5 level. At this spatial resolution, we assume that the incoming P2000 messages contain only postal code 5 information, by removing the last letter from the postal code as well as the house number. Each postal code 5 instance will be transformed back into object level instance, by means of the following procedure: A random subdistrict that lies in the postal code 5 area will be selected using data from the National Building Register. Subsequently, a random house number of the selected subdistrict will be chosen.

Then, steps 1 to 3 described in Section 4.2.1 can be taken. The last step, adding negative instances, could be taken by the following: select a random house as a negative instance that lies in another postal code 5 area (within the same district) than the positive rainwater damage instance.

Results

Spatial Resolution	Average accuracy	Average precision	Average recall
Object Level	0.5695	0.6001	0.4209
SubDistrict Level	0.5307	0.5317	0.5035
Postal Code 5 Level	0.5377	0.5355	0.5863
District Level	0.5514	0.5450	0.6263

Table 9: Cross-validated prediction results on various test sets with rainfall and height features in the normal sample range setting at various spatial resolutions of rainwater damage instances. Hereby, the evaluation metrics discussed in Section 4.5 are used.

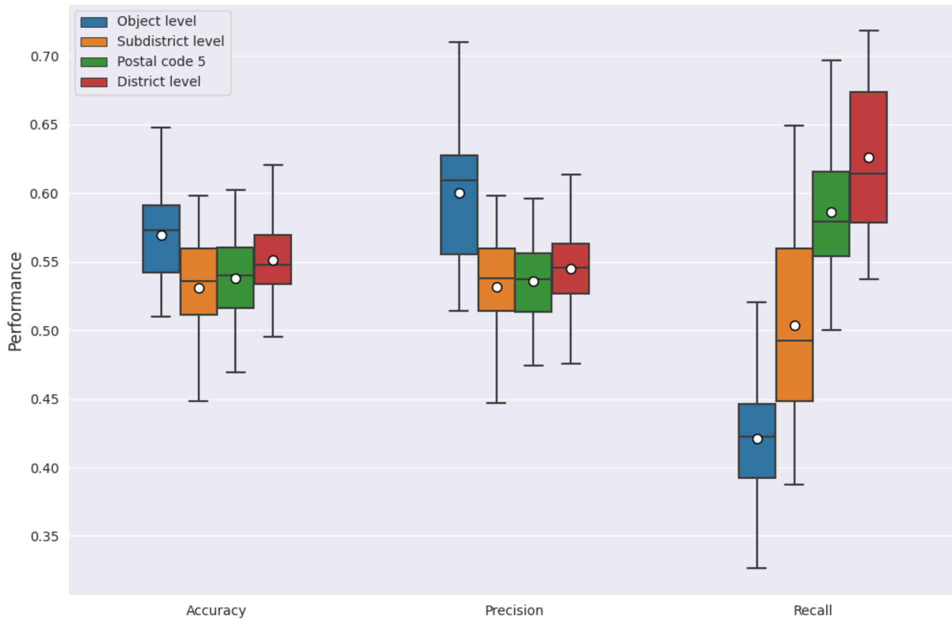


Figure 9: A box plot of the cross-validated prediction results of classifiers with postal code 5 level in the normal sample range setting. Hereby, height values and the past 3 hours rainfall are the predictive features.

Table 9 and Figure 9 present the following evaluation metrics for the random forest classifiers: accuracy, precision and recall as described in Section 4.5. In this setting, the classifiers at the object level have the highest accuracy score followed by classifiers at the district, postal code 5 and subdistrict levels. Our assumption that information loss will lead to a decrease in the performance of the classifiers seems therefore incorrect. However, there might be a logical reason for this. High differences in the amount of rainfall could enhance the performance of the classifiers at postal code 5 and district level because the positive and negative rainwater damage instances lie further away from each other.[2] So the models do not longer have to rely solely on height features. To further investigate this, we have conducted the second experiment setting where we increase the distance between positive and negative rainwater damage instances at all spatial resolutions: the district sample range setting. Hereby, each negative rainwater damage instance is sampled in another district from the positive rainwater damage instance, across all spatial resolutions.

[2] Note: the rainfall data is measured at 1 kilometre grid.

Spatial Resolution	Average accuracy	Average precision	Average recall
Object Level	0.5760	0.5870	0.5108
SubDistrict Level	0.5566	0.5519	0.6119
Postal Code 5 Level	0.5559	0.5446	0.6752
District Level	0.5514	0.5449	0.6263

Table 10: Cross-validated prediction results on various test sets with rainfall and height features in the district sample range setting at various spatial resolutions of rainwater damage instances. Hereby, the evaluation metrics discussed in Section 4.5 are used.

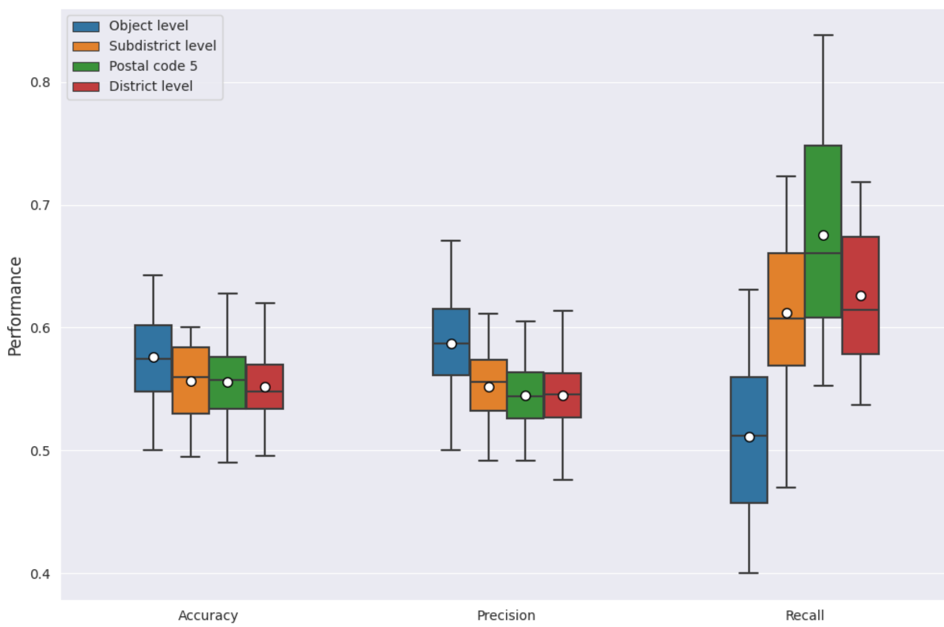


Figure 10: A box plot of the cross-validated prediction results of classifiers with postal code 5 level in the district sample range setting. Hereby, height values and the past 3 hours rainfall are the predictive features.

Table 10 and Figure 10 present the results of the district sample range setting with postal code 5 spatial resolution. Based on the results, we can conclude that gradual information loss leads to a decrease in the average accuracy performance of the random forest classifiers. In both settings, a trade-off in precision and recall can be noticed for classifiers at the object level. These results emphasise that the classifiers have the highest performance when complete information is presented in the water damage data.

B GitHub Repository

The GitHub repository of this project can be found at the following URL: `https://github.com/S127-Pi/rainwater_damage`