1 Berlin Topography

The city of Berlin is relatively flat and has mild terrain as shown in Figure 1.

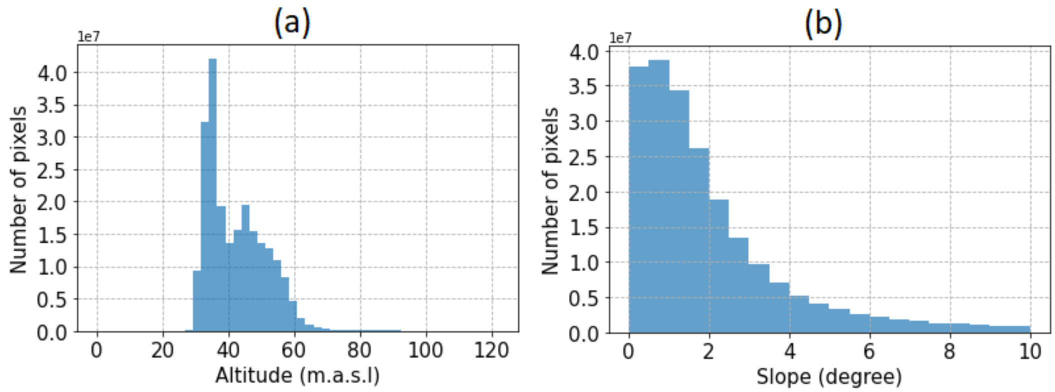

Figure 1: Berlin topographic characteristics: (a) altitude. (b) Slope.

2 Models hyper−parameters.

2.1 CNN and ANN models

Using two network architectures from the literature decreased the number of hyper−parameters for the CNN and ANN models. We used early stopping and dropout strategies to prevent overfitting. We examined several batch sizes and learning rates to get the best model performance for the CNN and ANN models. The optimal parameters for the two models are shown in Table 1 and 2 respectively.

Table 1: Best hyper−parameters values for CNN models

Spatial resolution	**Batch size**	**Learning rate**
30 m	512	0.0002633
10 m	512	0.0005725
5 m	128	0.00034863
2 m	128	0.00012819

2.2 Random forest models

The optimal hyper-parameters of for the RF models including the number of trees in the forest (n_estimators), the minimum number of samples to split an

Table 2: Best hyper−parameters values for ANN models

Spatial resolution	**Batch size**	**Learning rate**
30 m	512	0.0030955
10 m	512	0.00145
5 m	64	0.004737
2 m	512	0.002706

internal node (min_samples_split), the minimum number of samples required to be at a leaf node (min_samples_leaf), and the maximum depth of the tree (max_depth) are showing in Table 3

Table 3: Best hyper−parameters values for RF models

Spatial resolution	**n_estimators**	**min_samples_split**	**min_samples_leaf**	**max_depth**
30 m	200	5	2	90
10 m	1500	2	1	40
5 m	1000	2	1	50
2 m	500	5	1	None

2.3 Support vector machine

The optimal hyper-parameters of SVM including the penalty coefficient (C) and the radial basis function bandwidth (gamma) are shown in Table 4.

Table 4: Best hyper−parameters values for SVM models

Spatial resolution	**C**	**gamma**
30 m	30	10
10 m	20	10
5 m	10	10
2 m	100	10

2.4 Model Evaluation

Flood susceptibility mapping can be seen as a binary prediction problem (flood-prone vs. non flood-prone). Therefore, true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) can be calculated by comparing observations to model predictions.

Table 5: Performance indices

Index	Equation	Range
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	$0 <$ Accuracy < 1
F−score	$\frac{2TP}{2TP+FP+FN}$	$0 <$ F−score < 1

TP, TN, FP, and FN were used to compute the accuracy and F−score to evaluate the models' performance as shown in Table 5.

Figure 2 shows that all models perform very well on the training and testing datasets with performance indices all higher than 79 %. The RF models outperformed all other models at all spatial resolutions. and the RF model at 2 m resolution was superior for the present flood inventory and predictor data.

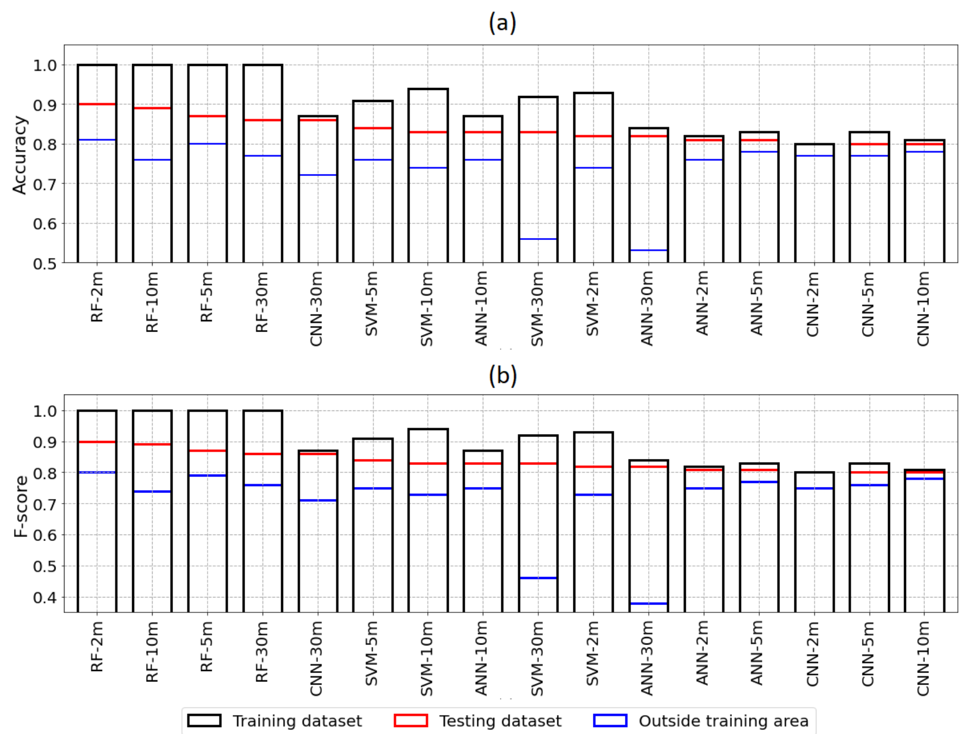

Figure 2: Calculated performance indices for the training dataset, testing dataset and the locations located outside training area, models are arranged horizontally in descending order according to their score for the testing dataset.

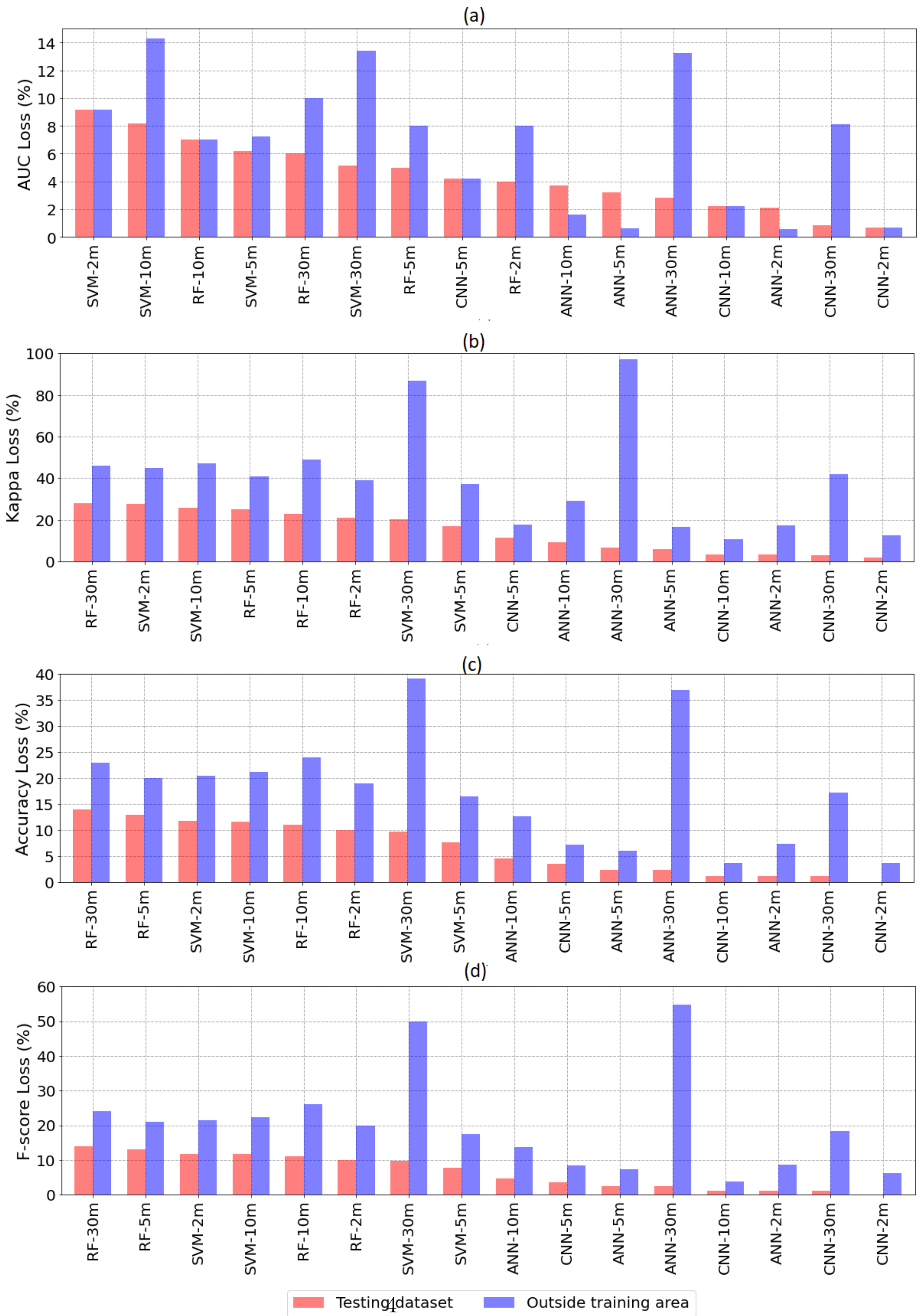

Figure 3: The loss in the performance indices score for the testing dataset and the locations outside the training area comparing to the score from the training dataset.