Flash flood predictions over the Eastern Mediterranean using artificial intelligence techniques with precipitable water vapor, pressure, and lightning data

Yuval Reuveni[1,2], Saed Asaly[3], and Lee-Ad Gottlieb[3]

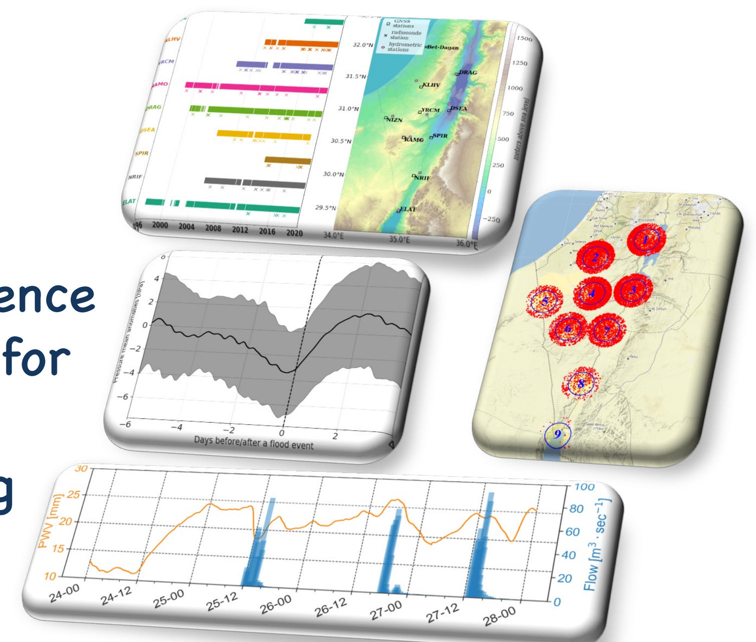
[1] *Department of Physics, Ariel University, Ariel, Israel*
[2] *Eastern R&D Center, Ariel University, Ariel, Israel*
[3] *Department of Computer Sciences, Ariel University, Ariel, Israel.*

Abstract

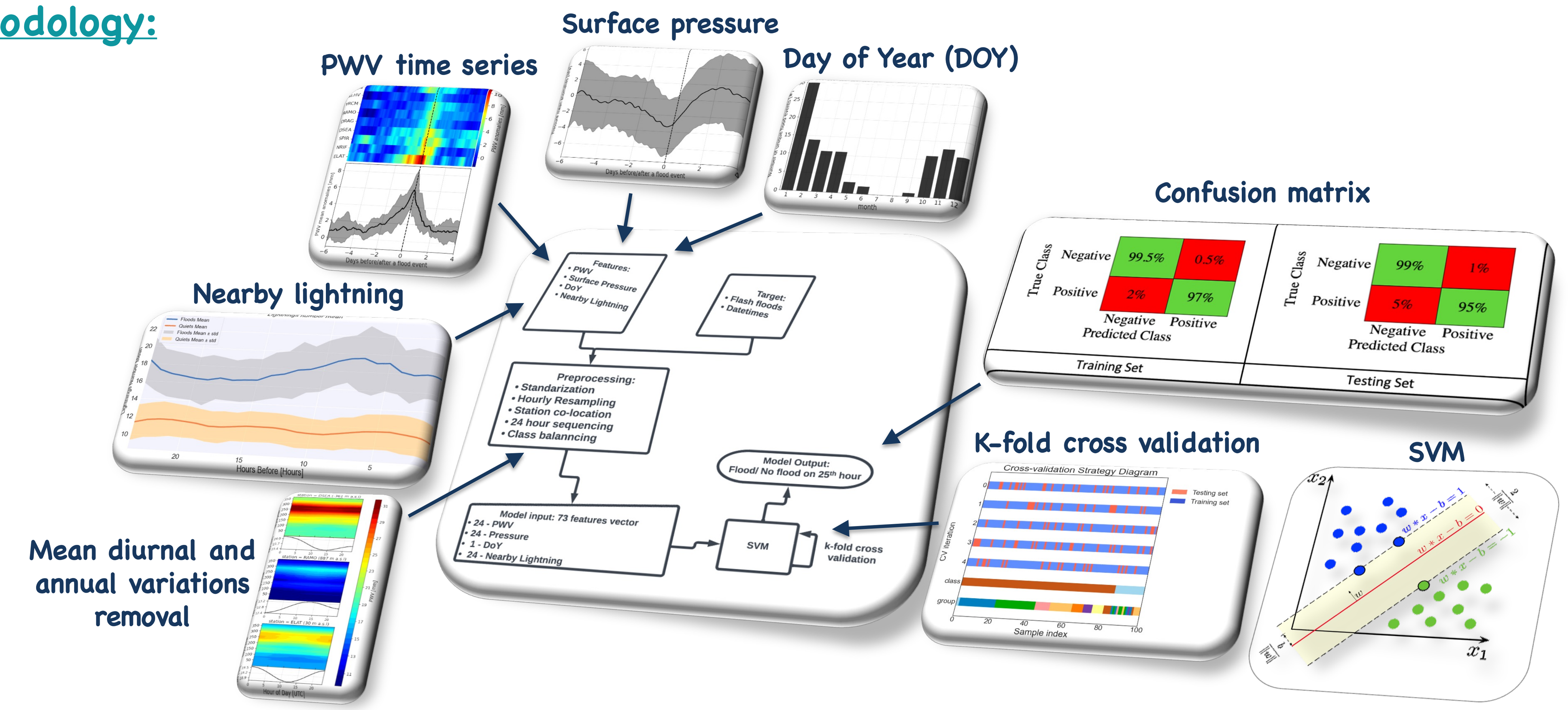
Here, we present a novel approach for improving flash flood predictions in the EM region using Support Vector Machines (SVMs) with a combination of precipitable water vapor (PWV) data, derived from ground-based global navigation satellite system (GNSS) receivers, along with surface pressure measurements, and nearby lightning occurrence data to predict flash floods in an arid region of the EM. The study found that integrating nearby lightning data with the other variables significantly improved the accuracy of flash flood prediction compared to using only PWV and surface pressure measurements. The results of the SVM model were validated using observed flash flood events, and the model was found to have a high predictive accuracy along with other high skill score metrics performances for the test set.

Data sources:

- Precipitable water vapor (PWV) derived from nine GNSS ground-based stations.
- Long-term hourly surface pressure measurements.
- Hydrometric station data which include the flood occurrence date times along with water level and water discharges for all recorded events.
- Lightning occurrence data from the WorldWide Lightning Location Network (WWLLN) and the Israel Lightning Detection Network (ILDN)

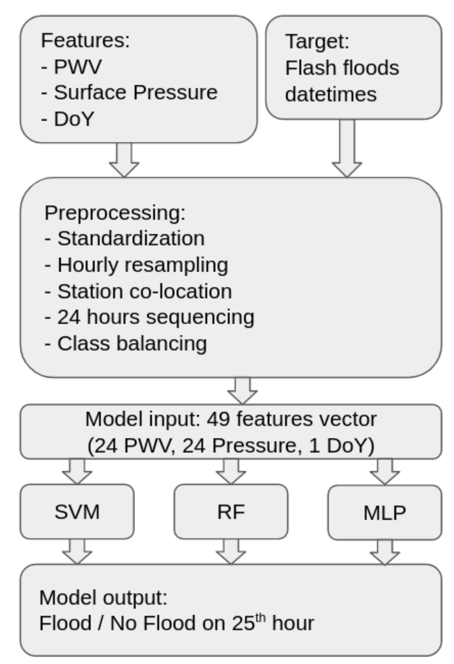


Methodology:



Methodology I:

Ziskin and Reuveni [2022]



Score metrics:

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{2}$$

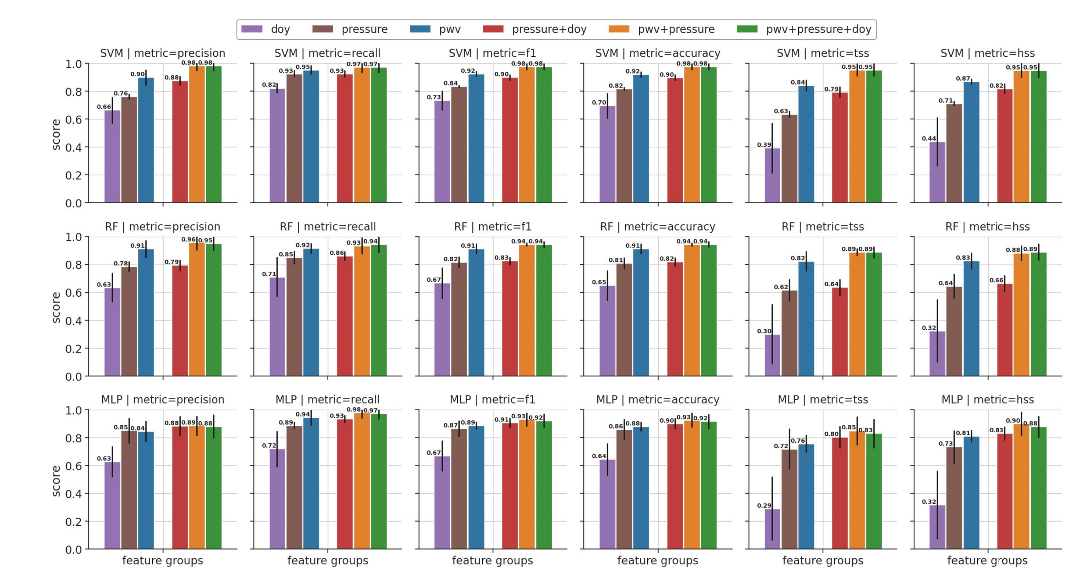
$$\text{HSS} = \frac{2\cdot[(\text{TP}\times\text{TN})-(\text{FN}\times\text{FP})]}{(\text{TP}+\text{FN})\cdot(\text{FN}+\text{TN})+(\text{TP}+\text{FN})\cdot(\text{FP}+\text{TN})} \tag{3}$$

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{P}+\text{N}} \tag{4}$$

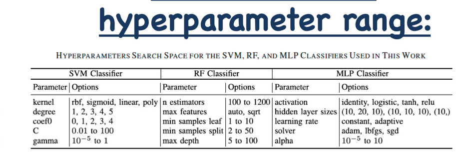
$$\text{TSS} = \frac{\text{TP}}{\text{TP}+\text{FN}} - \frac{\text{FP}}{\text{FP}+\text{TN}} \tag{5}$$

$$\text{F1} = 2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{6}$$

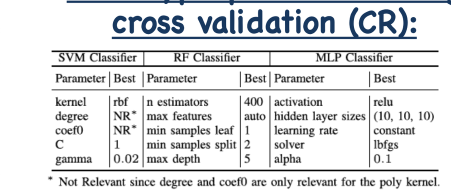
Mean test scores for the SVM, RF, and MLP classifiers (row) and for each metric (column):



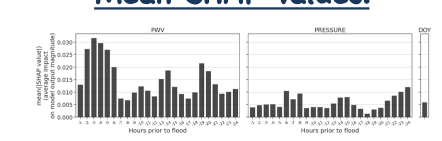
Grid search for hyperparameter range:



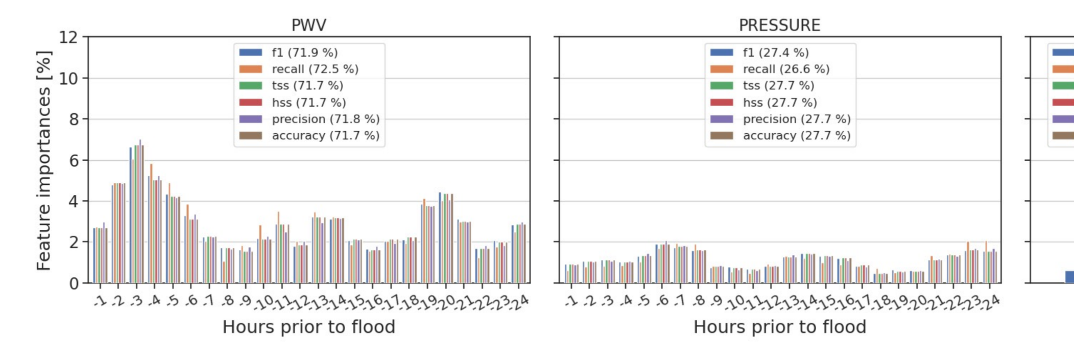
Best hyperparameter using cross validation (CR):



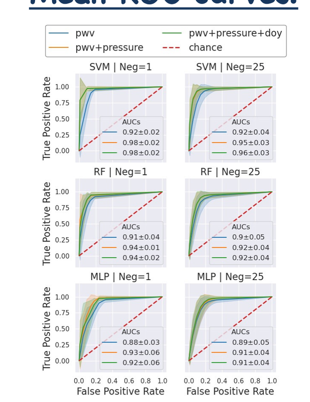
Mean SHAP values:



Feature importance for the PWV, surface pressure, and DoY features as run together in the RF classifier:

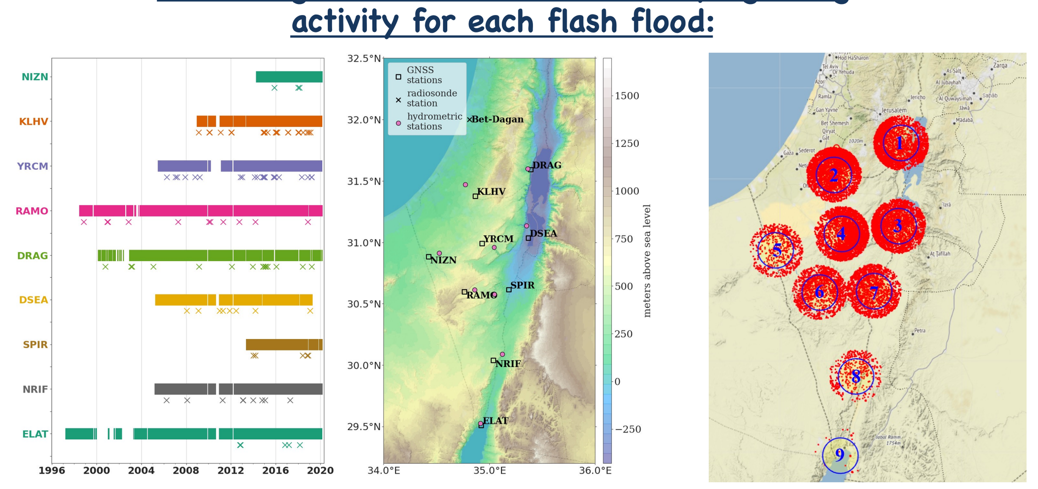


Mean ROC curves:

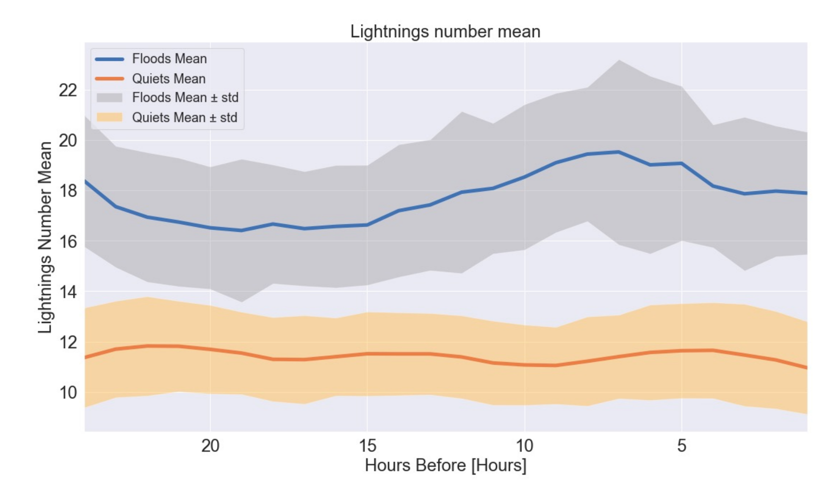


Methodology II:

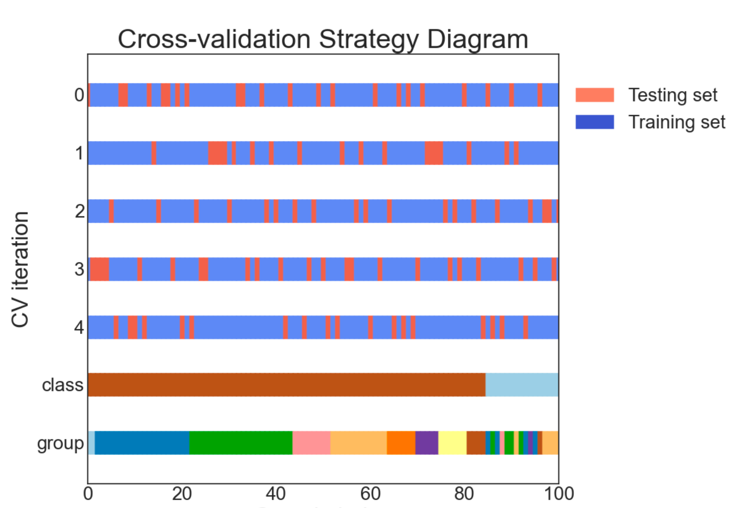
Extracting features related to nearby lightning activity for each flash flood:



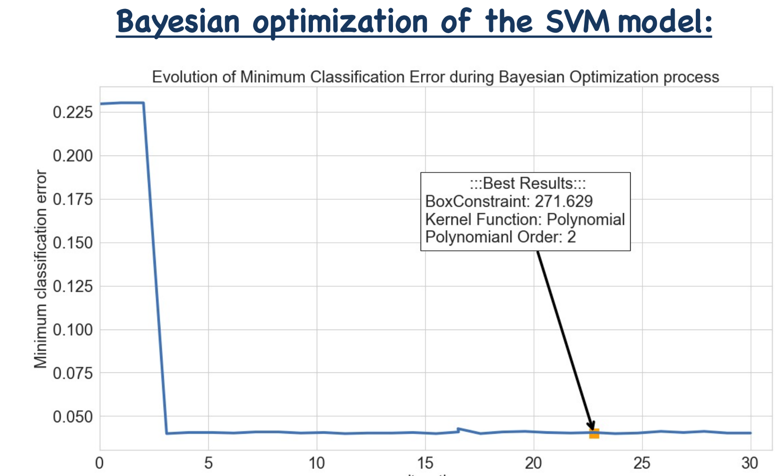
Constructing 24-hour vectors by integrating the number of lightnings which were nearby the GNSS and hydrometer station at a temporal resolution of 1-hour:



Cross validation performance with 5 subsets of the data using a randomized stratified sampling approach, allowing each iteration to randomly pick testing sets while still taking into account all the 9 stations (groups):

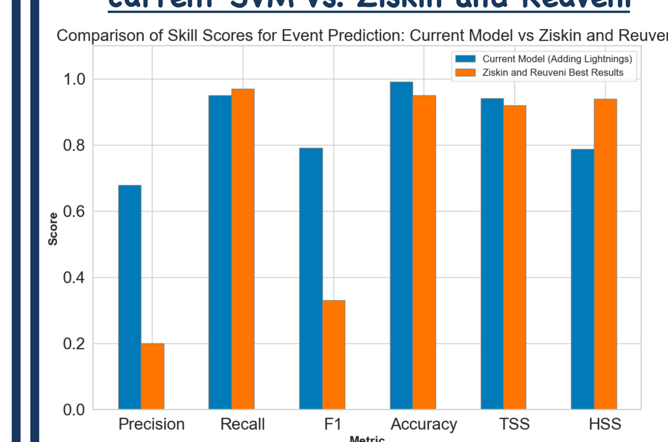


Bayesian optimization of the SVM model:

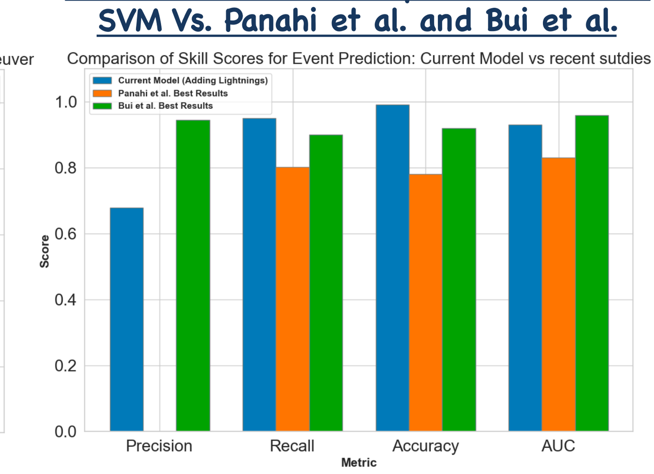


Results:

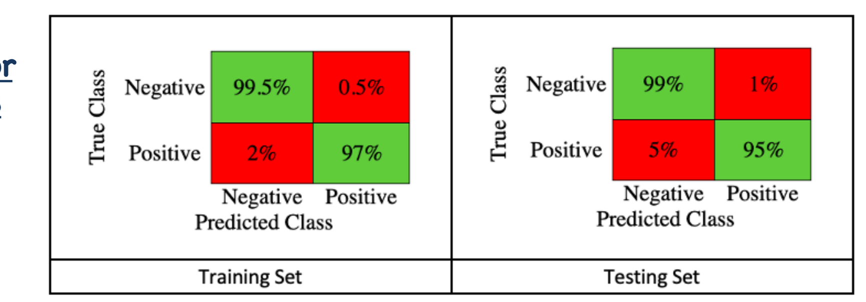
Skill score metrics compression: current SVM Vs. Ziskin and Reuveni



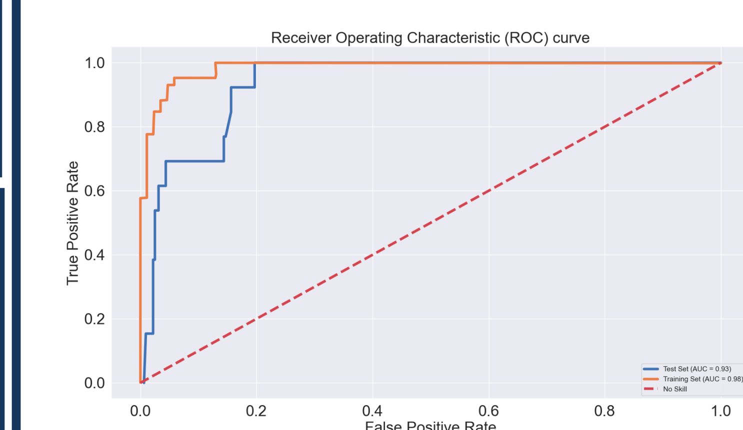
Skill score metrics compression: current SVM Vs. Panahi et al. and Bui et al.



The confusion matrix for the SVM model results extracted from the training set (left), and the test set (right):



ROC model curve obtained during the hyperparameters optimization process:



Conclusions:

Our results demonstrated that incorporating nearby lightning activity as an augmented feature, improved the performance of our model in capturing the correlation between atmospheric electricity characteristics and flash flood occurrence. This enhancement was evident in our model's precision and F1 metrics' performance on an imbalanced testing set, thus contributing to the development of a more accurate and reliable flash flood classification system.

References:

Ziv, S. Z., & Reuveni, Y. (2022). Flash Floods Prediction Using Precipitable Water Vapor Derived From GPS Tropospheric Path Delays Over the Eastern Mediterranean. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-17.

Asaly, S.; Gottlieb, L.; Yair, Y.; Price, C.; Reuveni, Y. (2023). Predicting Eastern Mediterranean Flash Floods using Support Vector Machines with precipitable water vapor, pressure, and lighting data. Remote Sensing. Under revision.

Panahi, M.; Jaafari, A.; Shirzadi, A.; Shahabi, H.; Rahmati, O.; Omidvar, E.; Lee, S.; Bui, D.T. (2021). Deep learning neural networks for spatially explicit prediction of flash flood probability. Geoscience Frontiers, 12, 101076.

Bui, D.T.; Hoang, N.D.; Mart.nez-.lvarez, F.; Ngo, P.T.T.; Hoa, P.V.; Pham, T.D.; Samui, P.; Costache, R. (2020) A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. Science of The Total Environment, 701, 134413.

Flash flood predictions over the Eastern Mediterranean using artificial intelligence techniques with precipitable water vapor, pressure, and lightning data

Yuval Reuveni[1,2], Saed Asaly[3], and Lee-Ad Gottlieb[3]

[1] *Department of Physics, Ariel University, Ariel, Israel*
[2] *Eastern R&D Center, Ariel University, Ariel, Israel*
[3] *Department of Computer Sciences, Ariel University, Ariel, Israel.*

Abstract

Here, we present a novel approach for improving flash flood predictions in the EM region using Support Vector Machines (SVMs) with a combination of precipitable water vapor (PWV) data, derived from ground-based global navigation satellite system (GNSS) receivers, along with surface pressure measurements, and nearby lightning occurrence data to predict flash floods in an arid region of the EM. The study found that integrating nearby lightning data with the other variables significantly improved the accuracy of flash flood prediction compared to using only PWV and surface pressure measurements. The results of the SVM model were validated using observed flash flood events, and the model was found to have a high predictive accuracy along with other high skill score metrics performances for the test set.

Background:

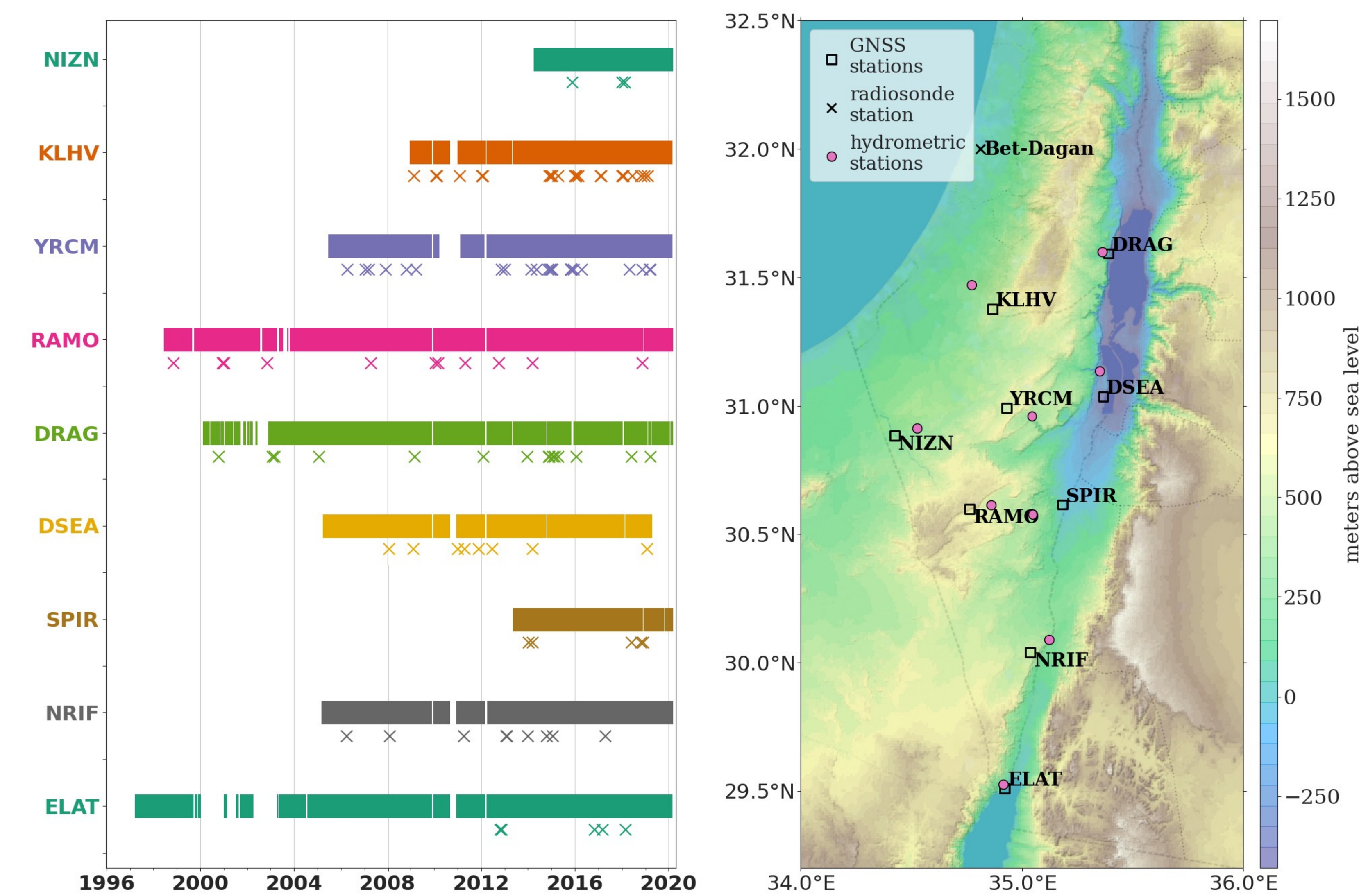
- A flash flood is a rapid and intense response of a drainage area to heavy rainfall events.
- The spatiotemporal distribution of rainfall is the most important factor (beside soil saturation and surface cover) for flash flood generation in the arid and semi-arid parts of the EM.
- A possible precursor to heavy rainfall events is the rise in tropospheric water vapor amount, which can be remotely sensed using ground-based global navigation satellite system (GNSS) stations.
- Heavy rainfall can lead to flood events and are often accompanied by an increase in nearby lightning activity.

Previuos work I:

- Ziskin and Reuveni (2022), examined 3 different machine learning methods, Support Vector Machine (SVM), Random Forest (RF), and Multilayered Perceptron (MLP), for binary classification task, which predicated whether a flash flood will occur given 24 hours of PWV, surface pressure, and a DOY feature:
- Using GNSS ground-based meteorology to monitor PWV before, during, and after heavy rainfall events. The PWV dataset used in this work has been derived from the SOI-APN GNSS ground receivers. We processed the daily RINEX files downloaded from the SOPAC/Garner GPS archive (http://garner.ucsd.edu/) using NASA's JPL GipsyX software. The daily RINEX processing is done using NASA's JPL GipsyX software via the PPP solution. From the position solution of the receiver, the ZTD can be extracted. We use a minimum cutoff elevation angle of 15°, GMF for the tropospheric model and ocean loading for all of the stations.

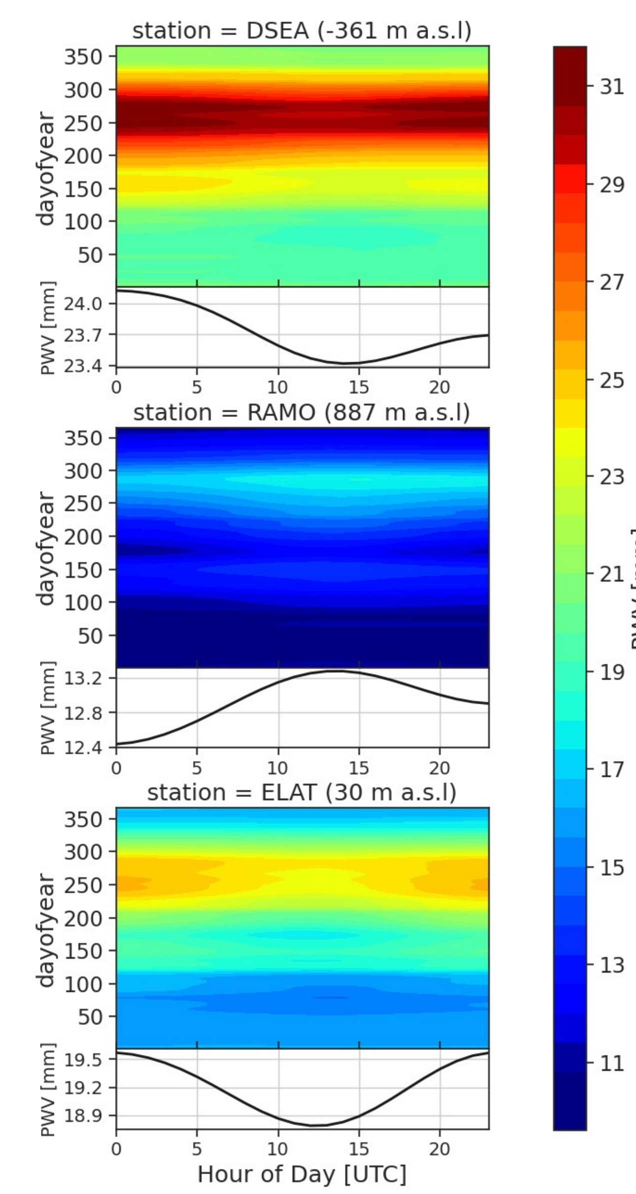
Previuos work II:

- The full parameter tree used in this work is available at the Github.com repository (https://github.com/ZiskinZiv/PW_from_GPS/blob/master/my_trees/ISROcnld/ISROcnld_0.tree). The processing has resulted in ZWD that was translated into PWV using the following formula: $PWV = \prod \times ZWD$. $\prod$ is the dimensionless constant of proportionality and is mainly the function of the atmospheric mean temperature. We used the Israeli Meteorological Service's (IMS) automated stations and radiosonde measurements in order to estimate the atmospheric mean temperature, Tm, relationship to the surface temperature, Ts , in the study area: Tm = 0.69Ts+82.



(Left) PWV data availability for each of the SOI-APN stations in the southern part of Israel. The flash floods' unique events are plotted with x's under each nearest GNSS station. (Right) SOI-APN stations (black squares), Bet-Dagan IMS station (black x), and the hydrometric stations (pink) plotted on a height-filled contour map of the study area.

- The final step in the PWV dataset preparation is the removal of the mean diurnal and annual variations. For each station, the resulting time series, which we call PWV anomalies, contains only the inter-daily variability:



PWV annual and diurnal climatology for (Top) DSEA, (Middle) RAMO, and (Bottom) ELAT stations. The diurnal annual mean is plotted under each filled contour panel

TABLE I

Geographical Coordinates, Altitude Above Sea Level, and the Names of the SOI-APN Stations in the Study Area

GNSS Station name	Station ID	Latitude [°N]	Longitude [°E]	Altitude [m a.s.l]
Nizana	NIZN	30.88	34.42	274
Kibutz Lahav	KLHV	31.38	34.87	498
Yerucham	YRCM	30.99	34.93	516
Mitzpe Ramon	RAMO	30.60	34.76	887
Metzoki dragot	DRAG	31.59	35.39	32
Dead-Sea Manufactories	DSEA	31.04	35.37	-361
Sapir	SPIR	30.61	35.18	12
Kibutz Neve Harif	NRIF	30.04	35.04	458
Eilat	ELAT	29.51	34.92	30

Flash flood predictions over the Eastern Mediterranean using artificial intelligence techniques with precipitable water vapor, pressure, and lightning data

Yuval Reuveni[1,2], Saed Asaly[3], and Lee-Ad Gottlieb[3]

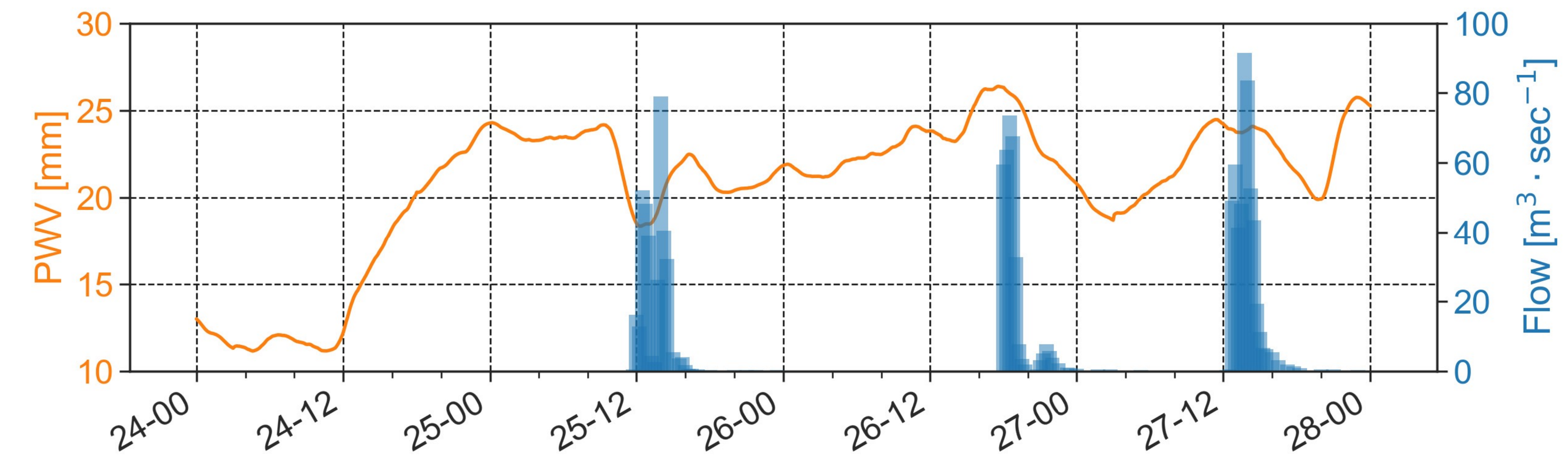
[1]Department of Physics, Ariel University, Ariel, Israel
[2]Eastern R&D Center, Ariel University, Ariel, Israel
[3]Department of Computer Sciences, Ariel University, Ariel, Israel.

EGU General Assembly 2023

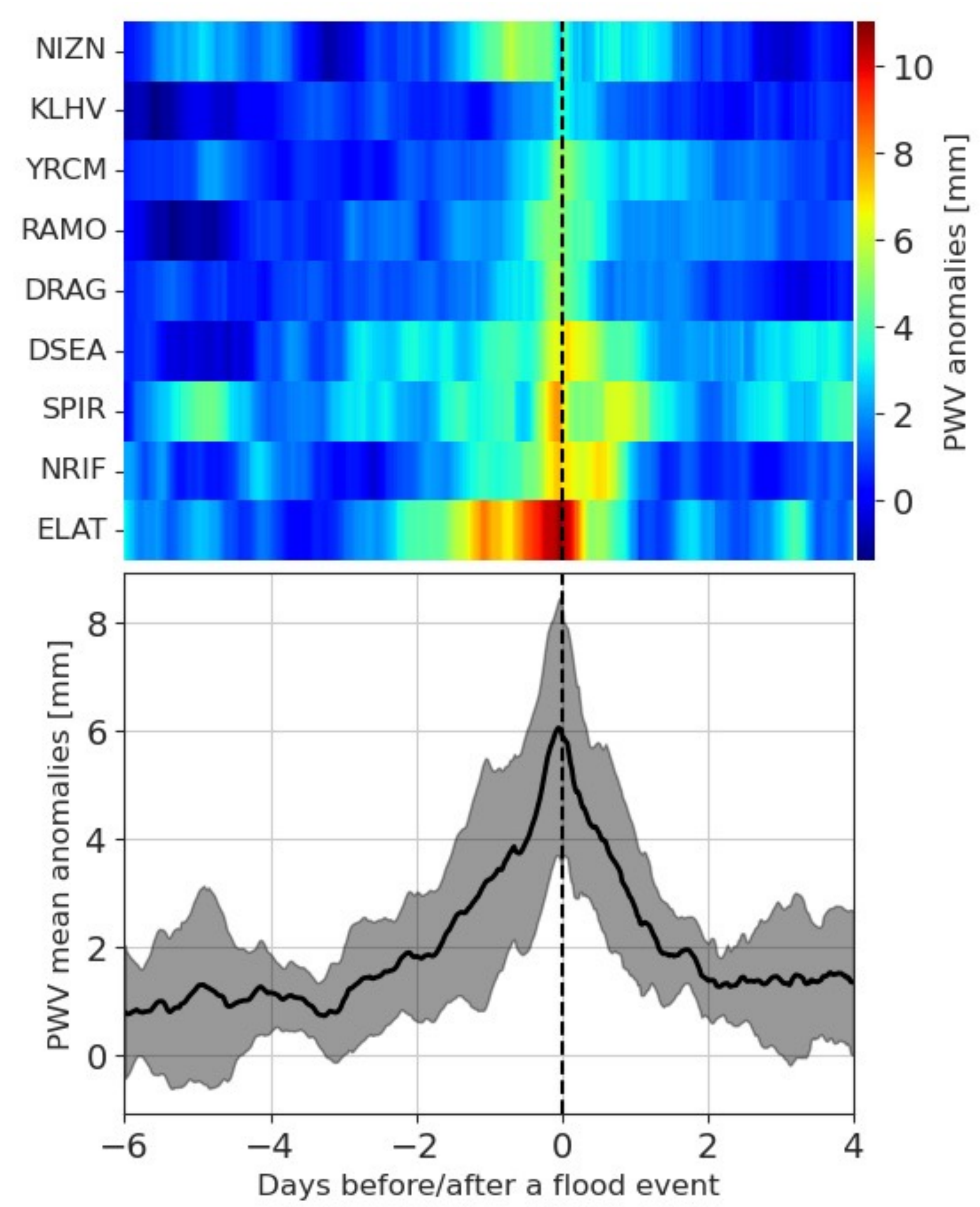
Previuos work II:

- The floods database has been received from the Israeli Water Authority (IWA, https://www.gov.il/en/departments/water_authority). The IWA manages and processes the measured data received from the hydrometric stations across Israel, which include the flood occurrence date times along with water level and water discharges for all recorded events. For each GNSS station, we searched for all available hydrometric stations located within a 15-km radius distance from the GNSS station location. We then selected the station with the highest amount of flood events, which we had the PWV data for, at least 24 h prior to the flood. Thus, we obtained an initial number of 151 flood events co-located with the respective GNSS stations.

PWV at Yerucham (YRCM) GNSS station superimposed on the water discharge (flow) at the Mamsheet hydrometric station located 12 km east of YRCM on April 24–27, 2018. Note the three major flash flood events on the 25th, the 26th, and the 27th. The PWV more than doubled during the second half of the 24th as a low-pressure system provided large quantities of moisture to the region



- In order to detect the effect that PWV has on flood events, we averaged the PWV anomalies six days prior and four days after a flood event. We repeated this step for all the GNSS stations and also averaged all the PWV anomalies stationwise:



(Top) PWV mean anomalies heatmap for the SOI-APN stations, presented in the map, with respect to a mean flood event. The average was calculated for various flood events (the rightmost column in Table II) per each station, from a total number of 151 events. (Bottom) Averaged PWV anomalies, along with its variability (indicated by the shaded gray strip), for the nine GNSS stations with respect to a ten-day time window around all the flood events (six days before and four days after the events, where the black dashed line is positioned at t = 0).

TABLE II
GEOGRAPHICAL COORDINATES, ALTITUDE ABOVE SEA LEVEL, AND THE NAMES OF THE HYDROMETRIC STATIONS ANALYZED IN THIS WORK

Hydrometric station name	Station ID	Latitude[°N]	Longitude[°E]	Altitude[m a.s.l]	Nearest GNSS station	Distance to GNSS station[km]	Flood events near GNSS station
Lavan - new nizana road	25191	30.91	34.53	251	NIZN	10	4
Shikma - Tel milcha	21105	31.47	34.77	202	KLHV	14	25
Mamsheet	55165	30.96	35.05	295	YRCM	12	25
Ramon	56140	30.61	34.86	480	RAMO	9	11
Draga	48125	31.60	35.37	-19	DRAG	3	15
Chiemar - down the cliff	48192	31.14	35.35	-320	DSEA	11	8
Nekrot - Top	56150	30.58	35.05	226	SPIR	14	5
Yaelon - Kibutz Yahel	60105	30.09	35.12	216	NRIF	10	9
Solomon - Eilat	60190	29.53	34.91	89	ELAT	2	5

Flash flood predictions over the Eastern Mediterranean using artificial intelligence techniques with precipitable water vapor, pressure, and lightning data

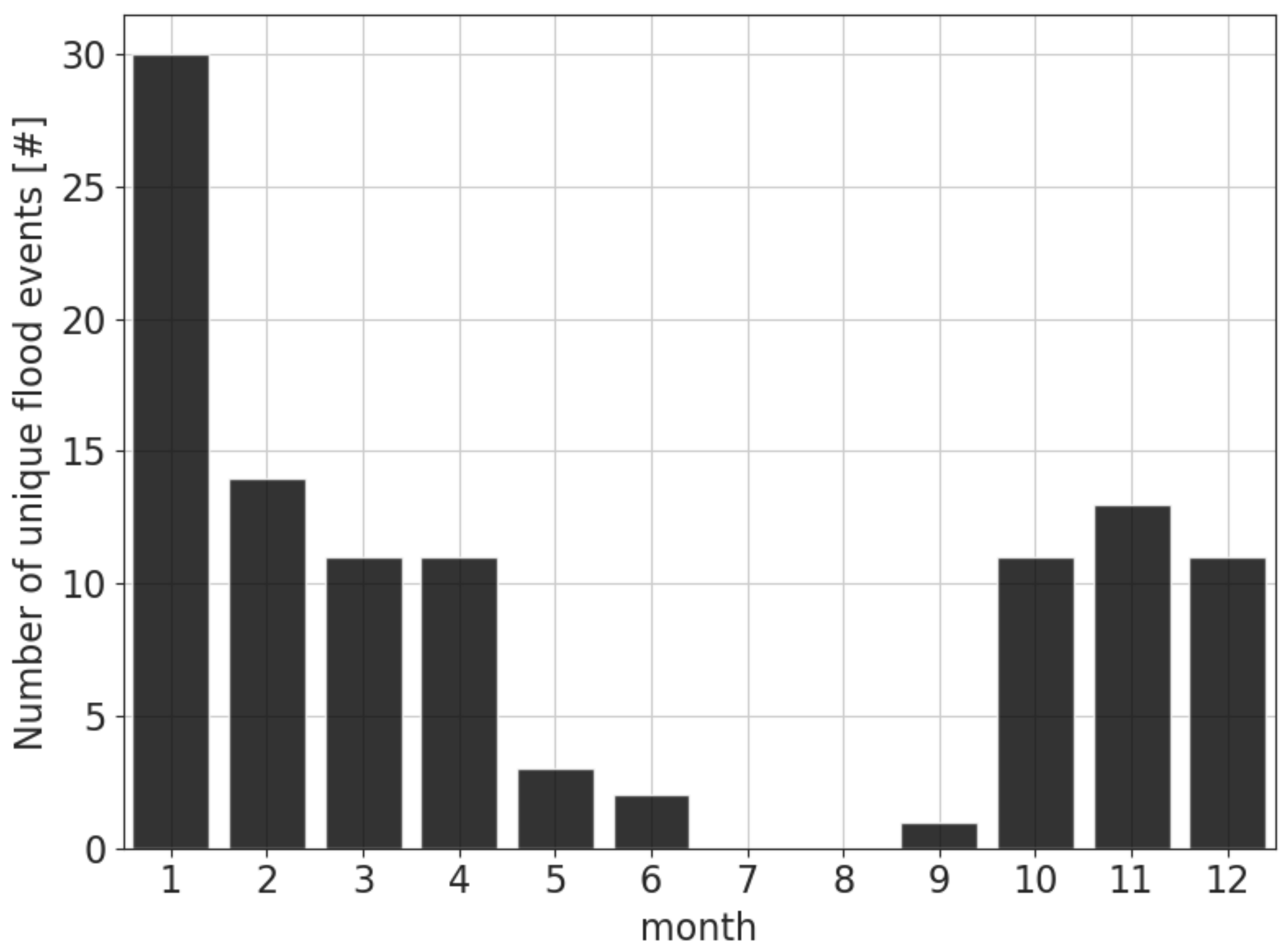
Yuval Reuveni[1,2], Saed Asaly[3], and Lee-Ad Gottlieb[3]

[1] Department of Physics, Ariel University, Ariel, Israel
[2] Eastern R&D Center, Ariel University, Ariel, Israel
[3] Department of Computer Sciences, Ariel University, Ariel, Israel.

Previuos work II:

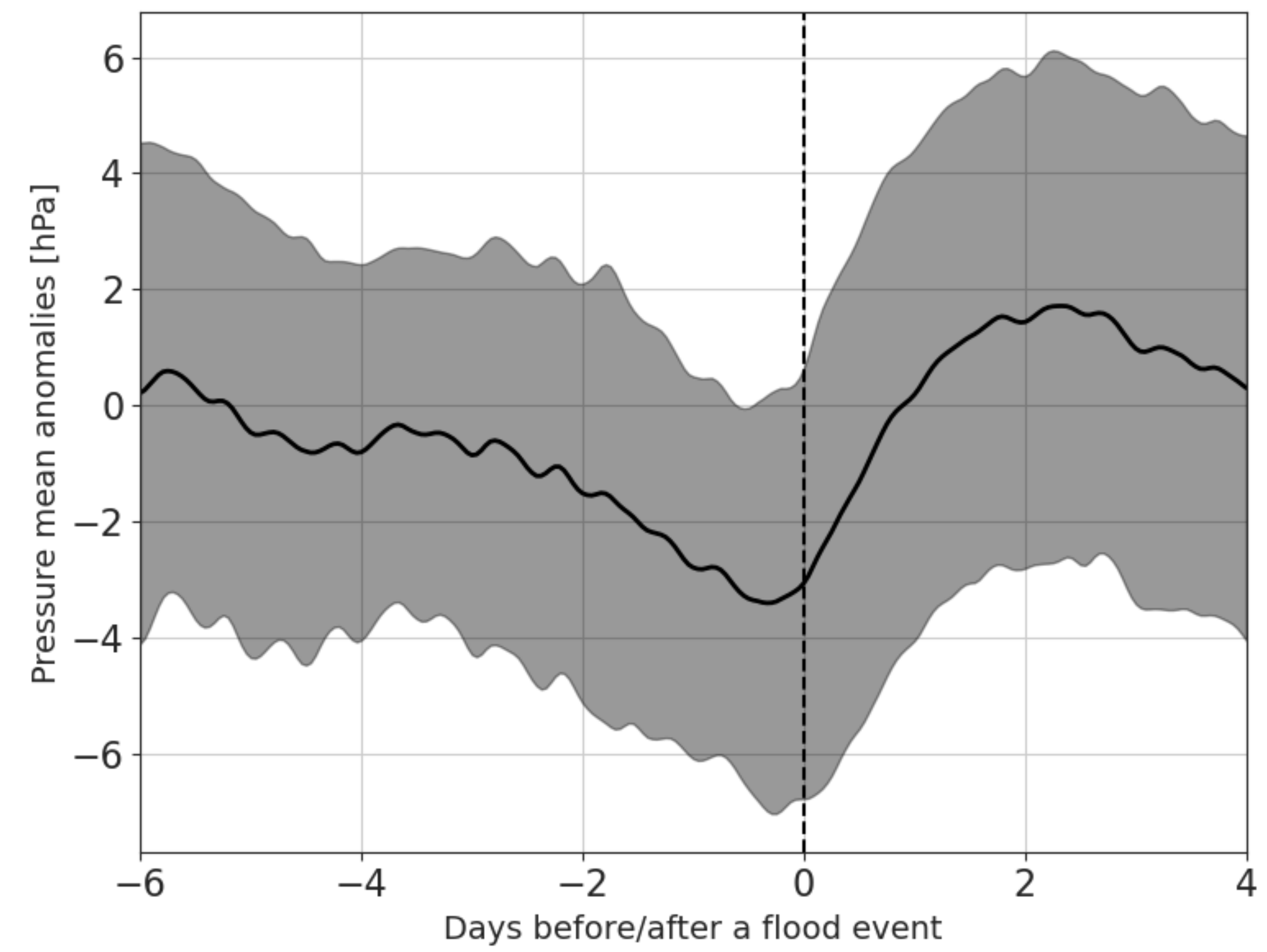
- Since our main approach to flash flood prediction is mostly data-driven, we decided to add more features with a goal of increasing our model's performance. In particular, we added long-term hourly surface pressure measurements from the Bet-Dagan IMS station (see map) and removed the diurnal and long-term climatology in the same manner as we did with the PWV data.

- We also added the Day of Year (DoY) information as a feature to our PWV and surface pressure features:



Number of flood events per month in the arid climate of southern Israel for events which we have PWV data for. It is clear that the most frequent month is January, with 30 events, while February-April and October-December have a mean of 11 events. May, June, and September have only a few events, while July and August have no flood events, as expected.

Station averaged pressure anomalies with respect to a mean flood event (black dashed line at x = 0). As expected, the pressure drops before a flood event, representing a low-pressure system that produces precipitation events. The minimum pressure values are found about 6-8 h prior to a flood event. However, the variability is quite higher than the PWV dataset.



Flash flood predictions over the Eastern Mediterranean using artificial intelligence techniques with precipitable water vapor, pressure, and lightning data

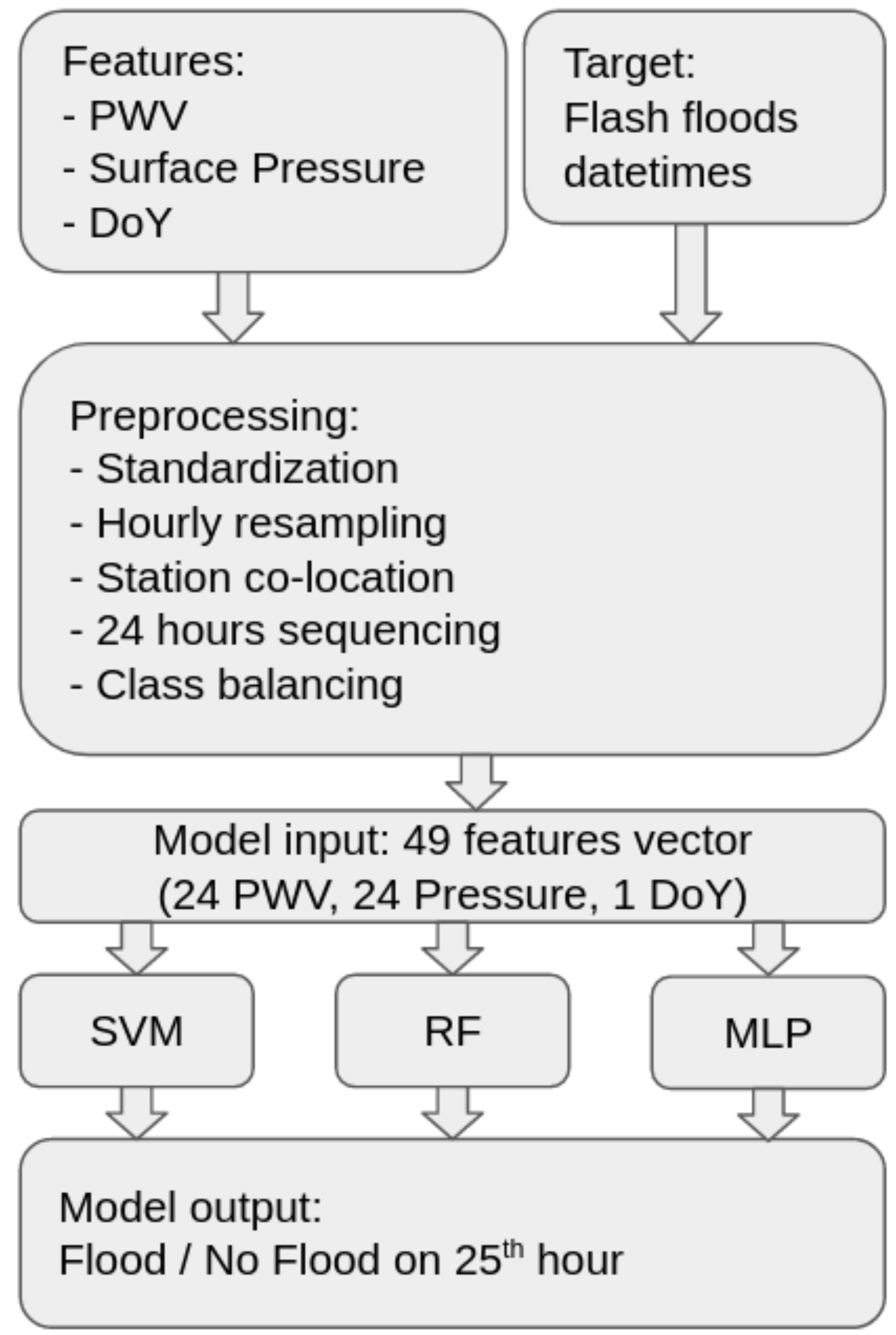
Yuval Reuveni[1,2], Saed Asaly[3], and Lee-Ad Gottlieb[3]

[1] Department of Physics, Ariel University, Ariel, Israel
[2] Eastern R&D Center, Ariel University, Ariel, Israel
[3] Department of Computer Sciences, Ariel University, Ariel, Israel.

Previuos work II:

ML Methodology:

- Preprocessing: Our data-driven approach to flood prediction considers a supervised learning task using binary classification. In particular, we ask the following question: given 24 h of PWV anomalies, surface pressure anomalies, and DoY, will there be a flood event in the following hour? When termed this way, we regard the PWV, surface pressure, and DoY data as features and the flood/nonflood events datetimes as the samples. Therefore, our preprocessing of the samples and features is given as follows.
  - First, we removed from the flood database close events that are overlapping within a 24-h window. The idea was to find unique flood events as much as possible, without losing too many samples. This step leaves us with 107 flood events from an original 151 GNSS co-located events. The flood events are the positive class in our classification task. We then continued with the positive features, i.e., PWV and surface pressure that are resampled to hourly means.
  - We then co-located each GNSS and hydrometric station and found 24 data points of PWV prior to each flood event. If half or more of the PWV data was missing, we dropped this event from our analysis. We used cubic interpolation to fill in the missing data points otherwise. We repeated this process with the surface pressure data, and however, in this case, we had only one surface pressure station (Bet Dagan) with the necessary data period and resolution. This step leaves us with 49 features (48 for PWV and pressure along with one for DoY). As for the negative class, we randomly searched for 24 h of PWV and pressure, which do not overlap the positive features, and we repeat this step only once for each flood event in each station, thus ensuring that the binary classification task is balanced. Our resulting matrix of features and samples is 214 (107 for each class) by 49. Finally, since two of our classifiers are sensitive to feature normalization, we use the standardized version of the PWV and surface pressure anomalies for all the classifiers (Standardized anomalies are the removal of the long-term monthly mean from a time series and dividing it by the long-term monthly standard deviation).
  - Our main goal is to use supervised learning classifiers in order to predict flash floods using PWV as the main input. Accordingly, we chose three common types of ML models: SVM, RF, and multilayered perceptron (MLP). All the models were implemented using the Scikit-Learn Python package. **The SVM classifier** utilizes a linear hyperplane to separate each sample class. Using the kernel trick, the hyperplane is transformed into a higher dimension, which gives the SVM more flexibility; however, the cost is a larger generalization error. **The RF classifier** is a metaclassifier, which uses a number of decision trees on randomized selections of subset of features. The final output is produced by averaging all the individual decision tree classifiers. **The MLP classifier** is a neural network algorithm, which includes multilayered nodes with weights. Typically, the network architecture includes an input layer, any number of hidden layer, and an output layer where each layer's nodes are connected via activation functions (a so-called feedforward propagation). During the learning process, the weights are reevaluated using the backpropagation iterative algorithm [54] in order to decrease the cost function.



Station Main ML methodology block diagram. The features are the PWV, surface pressure, and Doy, where the target is the flash floods datetimes. Preprocesing involves standardizing the PWV and surface pressure measurements (Standardized anomalies are the removal of the long-term monthly mean from a time series and dividing it by the long-term monthly standard deviation), hourly resampling them, and colocating the GNSS and hydrometric stations. Finally, 24-h sequences are generated with class balancing. In the learning process, three general types of ML classifier models are optimized using Cross Validation: MLP, SVM, and RF. The final output of each model is whether or not a flash flood will occur in the 25th hour.

Flash flood predictions over the Eastern Mediterranean using artificial intelligence techniques with precipitable water vapor, pressure, and lightning data

Yuval Reuveni[1,2], Saed Asaly[3], and Lee-Ad Gottlieb[3]

[1] Department of Physics, Ariel University, Ariel, Israel
[2] Eastern R&D Center, Ariel University, Ariel, Israel
[3] Department of Computer Sciences, Ariel University, Ariel, Israel.

EGU General Assembly 2023

Previuos work II:

ML Methodology:

- score Metrics: We use six different metrics to evaluate the models' performance [55]. These metrics are: precision, recall, F1, accuracy, Heidke skill score (HSS), and true skill statistics (TSS). These metrics are a combination of the four possible outcomes of our classifier.
  1) True positive (TP) is the correct prediction of a flood event.
  2) True negative (TN) is the correct prediction of a no-flood situation.
  3) False positive (FP or type I error or false alarm) is when the classifier predicts a flood but there was not any.
  4) False negative (FN or type II error or simply miss) is when the classifier does not predict a flood but a flood occurs, hence the miss.

- The fallout or false positive rate (FPR), measures the probability of false alarm (FPs). The precision or positive predictive value measures the ability of the classifier not to produce false alarms. The recall also known as true positive rate (TPR), sensitivity, or hit rate measures how successful the classifier is in predicting the positive class without missing (FN). Precision and recall are always at tension with each other, where improving recall reduces the precision and vice versa. One way of dealing with this issue is to use the F1 score, which is the harmonic mean of the precision and recall. The accuracy score quantifies how well a classification test correctly identifies or excludes a condition (i.e., whether it is a TP or TN). The TSS compares the probability of the true prediction, to the probability of false prediction or simply recall minus the fallout. Thus, a TSS no skill score is 0, while -1 means that the prediction labels should be reversed. The HSS, which is often used in weather and solar events prediction, quantifies the fractional improvement of the prediction accuracy relative to some set of control or reference predictions. It is normalized by the total range of possible improvement over the standard (i.e., it can be compared with different datasets). A perfect HSS score is 1, and a no skill score is 0, while an infinitely negative score is possible, suggesting that the prediction is worse than the reference prediction. Another widely used performance measurement visualization method is the receiver operating characteristics (ROC) curve, which illustrates the diagnostic ability of a binary classifier as its classification threshold is varied. The ROC curve is actually the recall or TPR plotted versus the fallout or FPR where, ideally, the TPR is maximized, while the FPR is minimized. The area under the ROC curve (ROC-AUC) can be used as a score metric where a no skill score is 0.5, while a perfect score is 1.

- k-fold nested cross-validation: The models were tested with a technique called k-fold nested cross-validation, which is a technique used to evaluate the performance of machine learning models by dividing a dataset into multiple subsets, or "folds," and iteratively training and testing the model on different combinations of these folds. In K-fold Nested Cross-Validation, the process is repeated multiple times, with each iteration splitting the data into different folds and using different subsets of the data for training and testing. This approach is used to help reduce overfitting and ensure that the model is not biased towards a particular subset of the data.
  The "nested" aspect of this technique refers to the fact that it involves two levels of cross-validation. The outer loop of the process uses K-fold cross-validation to split the data into training and testing sets, while the inner loop is used to perform hyperparameter tuning on the model. The hyperparameter optimization process has been done using a technique called grid-search optimization, which is basically a technique used to find the best set of hyperparameters for a machine learning model. Hyperparameters are values that are set before training the model and can affect how well the model performs on new data.

  In Grid Search Optimization, a range of values is specified for each hyperparameter, and the model is trained and evaluated on each combination of these values. This is done by creating a "grid" of all possible hyperparameter values and evaluating the model on each point on the grid.

Score metrics:

$$\text{Fallout} = \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

$$\text{Precision} = \text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

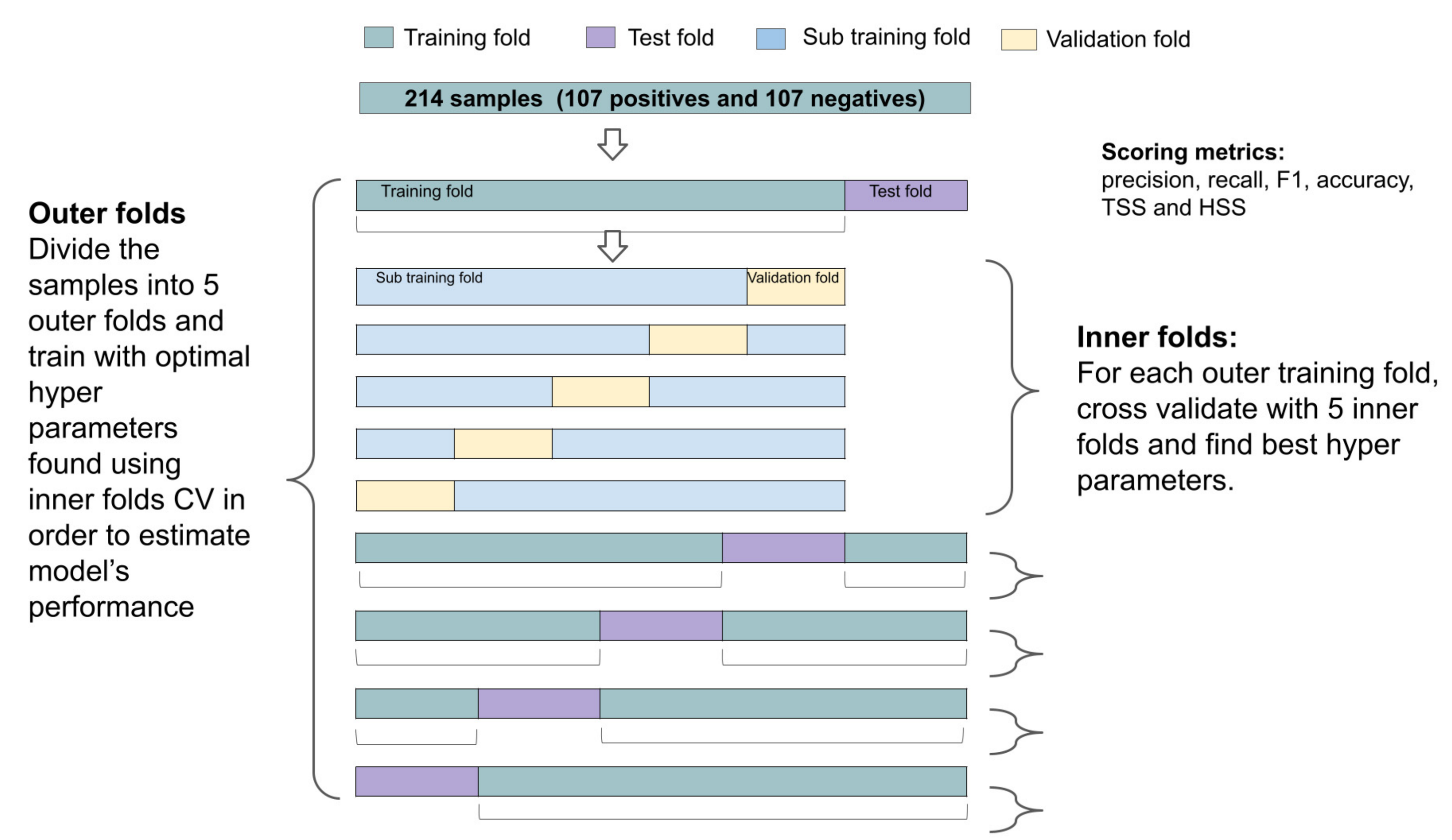
$$\text{Recall} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{TSS} = \frac{\text{TP}}{\text{TP} + \text{FN}} - \frac{\text{FP}}{\text{FP} + \text{TN}} = \text{Recall} - \text{Fallout}$$

$$\text{HSS} = \frac{2 \times [\text{TP} \times \text{TN} - \text{FN} \times \text{FP}]}{(\text{TP} + \text{FN}) \times (\text{FN} + \text{TN}) + (\text{TP} + \text{FN}) \times (\text{FP} + \text{TN})}.$$



Flash flood predictions over the Eastern Mediterranean using artificial intelligence techniques with precipitable water vapor, pressure, and lightning data

ARIEL UNIVERSITY

EASTERN R&D CENTER

EGU General Assembly 2023

Yuval Reuveni[1,2], Saed Asaly[3], and Lee-Ad Gottlieb[3]

[1]*Department of Physics, Ariel University, Ariel, Israel*
[2]*Eastern R&D Center, Ariel University, Ariel, Israel*
[3]*Department of Computer Sciences, Ariel University, Ariel, Israel.*

Previuos work II:

ML Methodology:

- Permutation Test: We also subject our classifiers to the permutation test for labeled data. This test, which has been extensively used in the field of computational biology, aims to address the following question: does the classifier detect a significant class structure, i.e., a real connection between the data and the class labels? We use a standard fivefold CV to estimate a null distribution by permuting the labels in the data and produce a "true" score without the permutations. The experimental p-value from these tests is calculated as follows: $p-\text{value} = (S + 1)/(n_{\text{permutations}} + 1)$, where S is the number of permutations whose score the "true" score. Since ideally, S should be 0, the best possible p-value is $1/(n_{\text{permutations}} + 1)$, and since we use 100 permutations, it is 1/101 = 0.0099, while the worst p-value is when $S = n_{\text{permutations}}$, i.e., p-value = 1.0.

- Imbalanced Dataset Test: Since flash floods are very rare events, we thus require a more realistic scenario for testing our classifier, which is trained with a balanced dataset. Therefore, we need to generate more negative samples from the PWV/pressure time series. As a rough estimate, we divide the number of the total flash flood events (≈100) with the total number of days of the largest time series (RAMO: ≈ 7500 days or ≈ 20.5 years) and reach a ratio of 1 flash flood event in 75 days or 1.3% positive ratio. Thus, we need to produce negative samples for each station that is complete (24 h) and do not coincide with a positive event. Unfortunately, with these constraints, we were able to find only 25 negative samples per a positive one or 4% positive ratio that is three times more frequent than the rough estimate. Nevertheless, we can use a specific data split in order to overcome this obstacle. The testing procedure for the imbalanced dataset is given as follows:
  1) For each ML model, we train our classifiers with 66.66% of the balanced training set (71 positives and 71 negatives).
  2) We evaluate the classifiers with the remaining 33.33% of the balanced dataset concatenated with all the remaining negative samples produced (36 positives and 2639 negatives) to receive a positive ratio of 1:73.3 or 1.36%, which is very close to our estimate.
  3) We repeat the evaluation for each of the score metrics.

Grid search for hyperparameter range:

HYPERPARAMETERS SEARCH SPACE FOR THE SVM, RF, AND MLP CLASSIFIERS USED IN THIS WORK

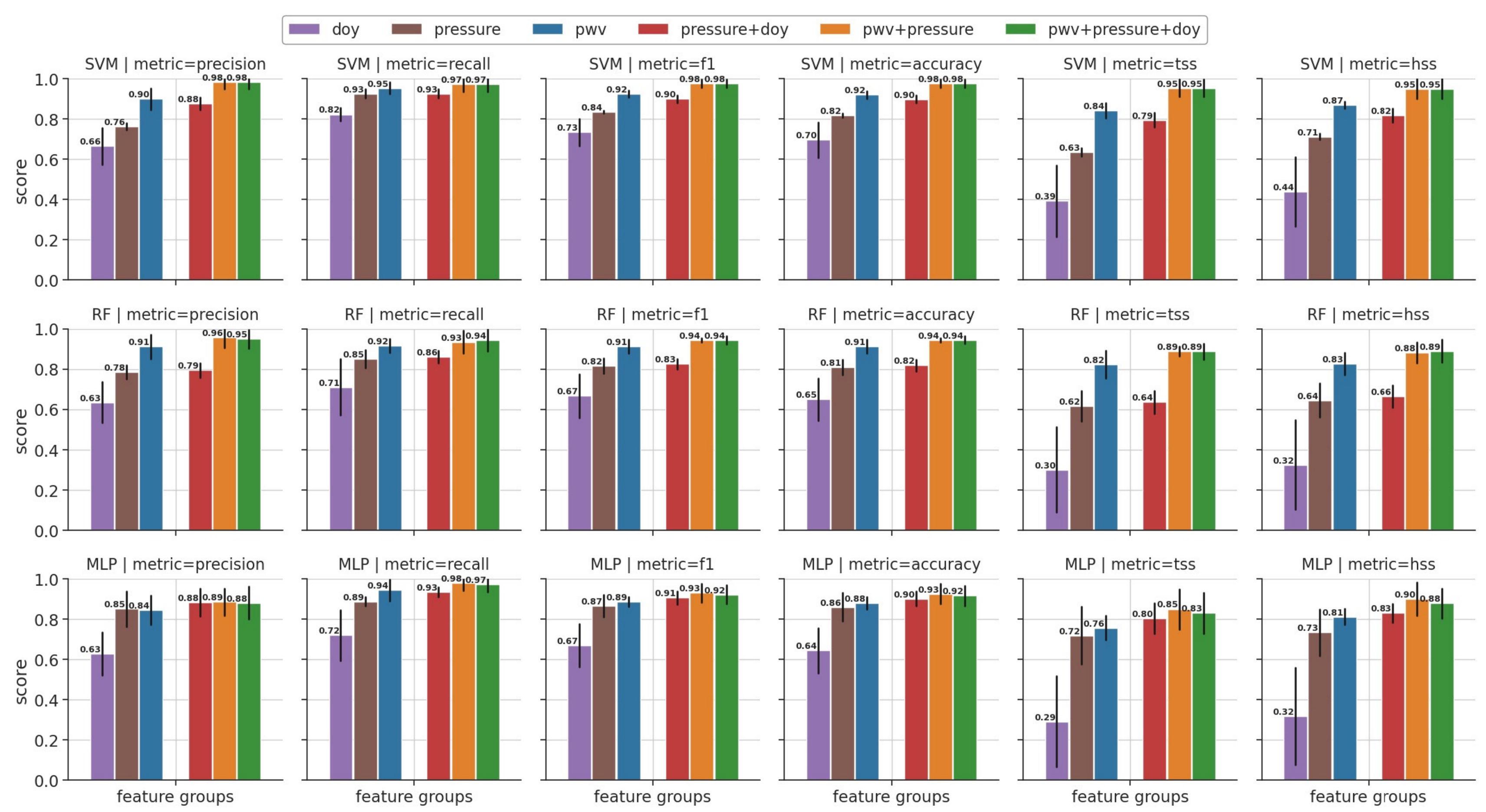
SVM Classifier		RF Classifier		MLP Classifier	
Parameter	Options	Parameter	Options	Parameter	Options
kernel	rbf, sigmoid, linear, poly	n estimators	100 to 1200	activation	identity, logistic, tanh, relu
degree	1, 2, 3, 4, 5	max features	auto, sqrt	hidden layer sizes	(10, 20, 10), (10, 10, 10), (10,)
coef0	0, 1, 2, 3, 4	min samples leaf	1 to 10	learning rate	constant, adaptive
C	0.01 to 100	min samples split	2 to 50	solver	adam, lbfgs, sgd
gamma	10^{-5} to 1	max depth	5 to 100	alpha	10^{-5} to 10

Best hyperparameter using cross validation (CR):

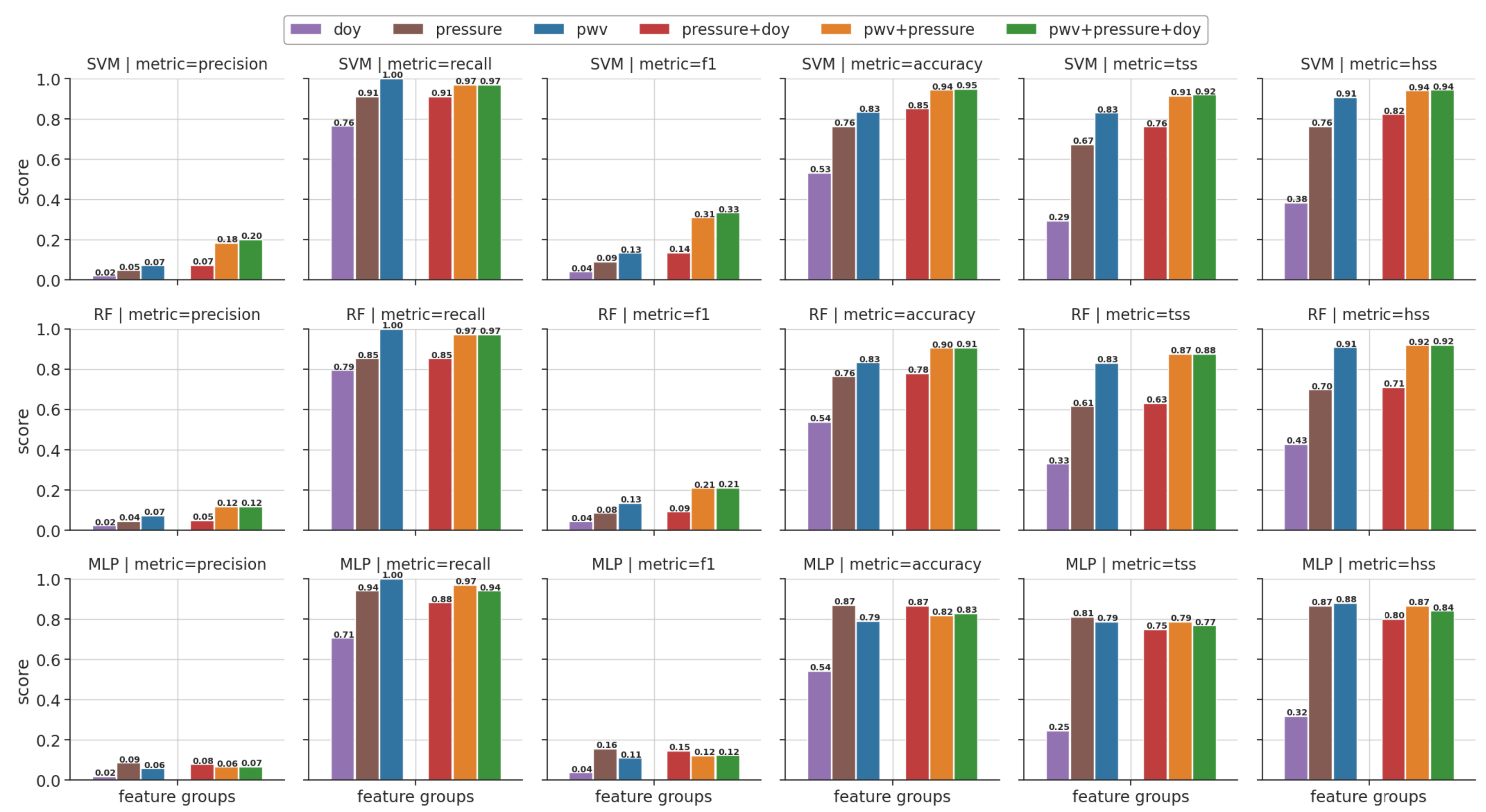
SVM Classifier		RF Classifier		MLP Classifier	
Parameter	Best	Parameter	Best	Parameter	Best
kernel	rbf	n estimators	400	activation	relu
degree	NR*	max features	auto	hidden layer sizes	(10, 10, 10)
coef0	NR*	min samples leaf	1	learning rate	constant
C	1	min samples split	2	solver	lbfgs
gamma	0.02	max depth	5	alpha	0.1

* Not Relevant since degree and coef0 are only relevant for the poly kernel.

Mean test scores for the SVM, RF, and MLP classifiers (row) and for each metric (column). The feature groups consist of DoY (purple), surface pressure (brown), PWV (blue), surface pressure and DoY (red), PWV and surface pressure (orange), and all three together (green). The mean scores are indicated to the top left of each bar and the SD of five data splits is represented by the error bar length.



Imbalanced dataset test scores for the SVM, RF, and MLP classifiers (row) and for each metric (column).



Flash flood predictions over the Eastern Mediterranean using artificial intelligence techniques with precipitable water vapor, pressure, and lightning data

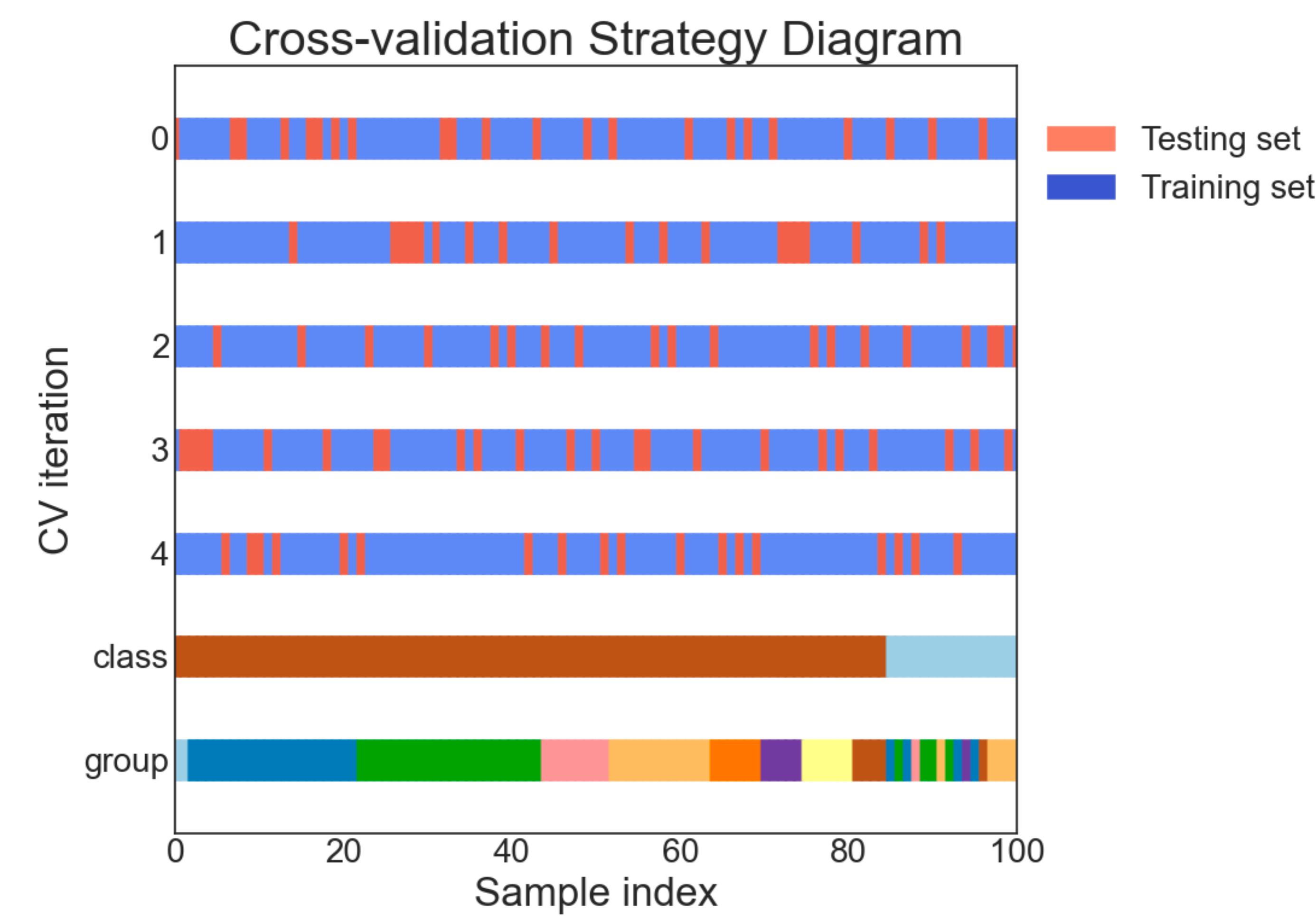
Yuval Reuveni[1,2], Saed Asaly[3], and Lee-Ad Gottlieb[3]

[1]*Department of Physics, Ariel University, Ariel, Israel*
[2]*Eastern R&D Center, Ariel University, Ariel, Israel*
[3]*Department of Computer Sciences, Ariel University, Ariel, Israel.*

Current work:

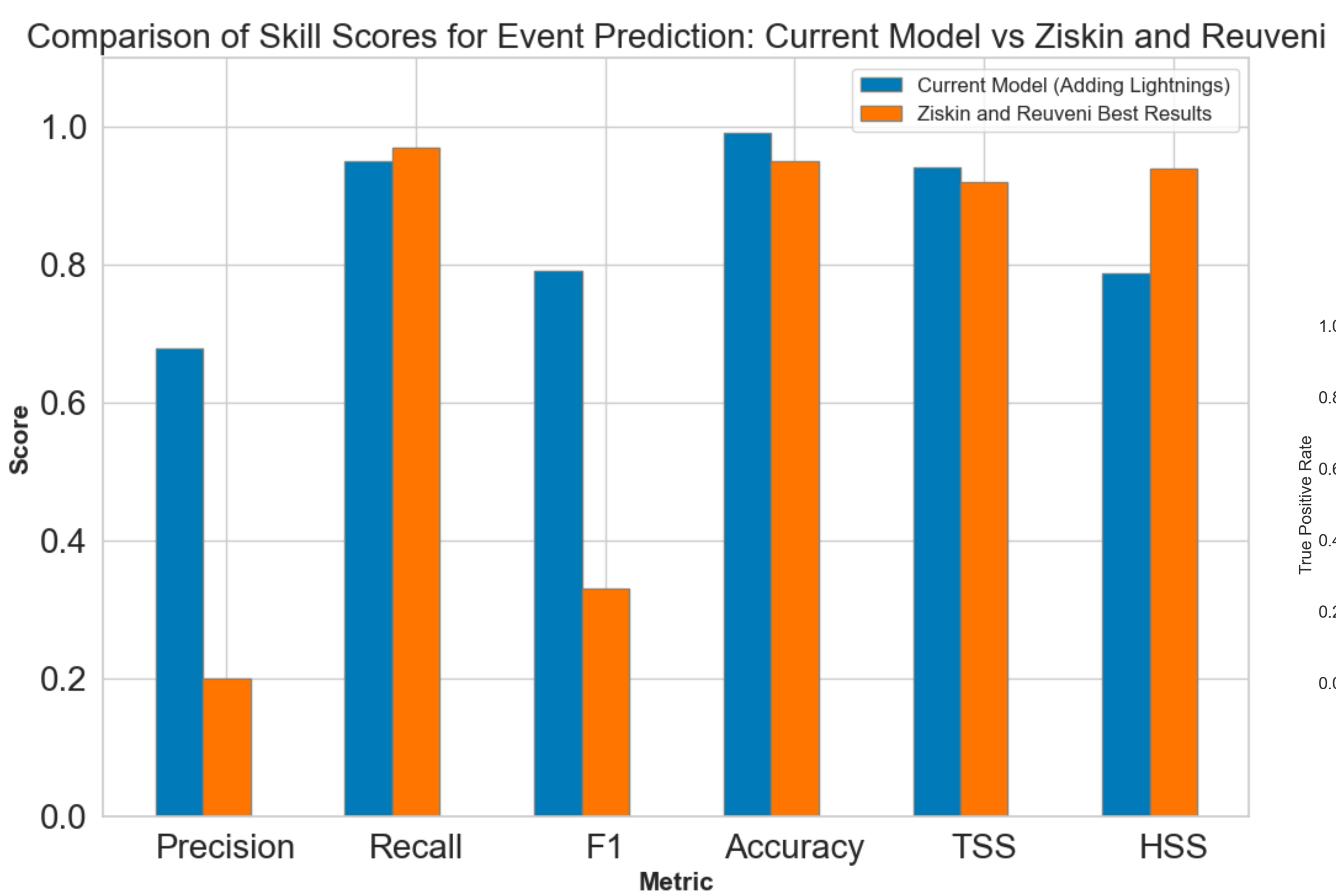
ML Methodology:

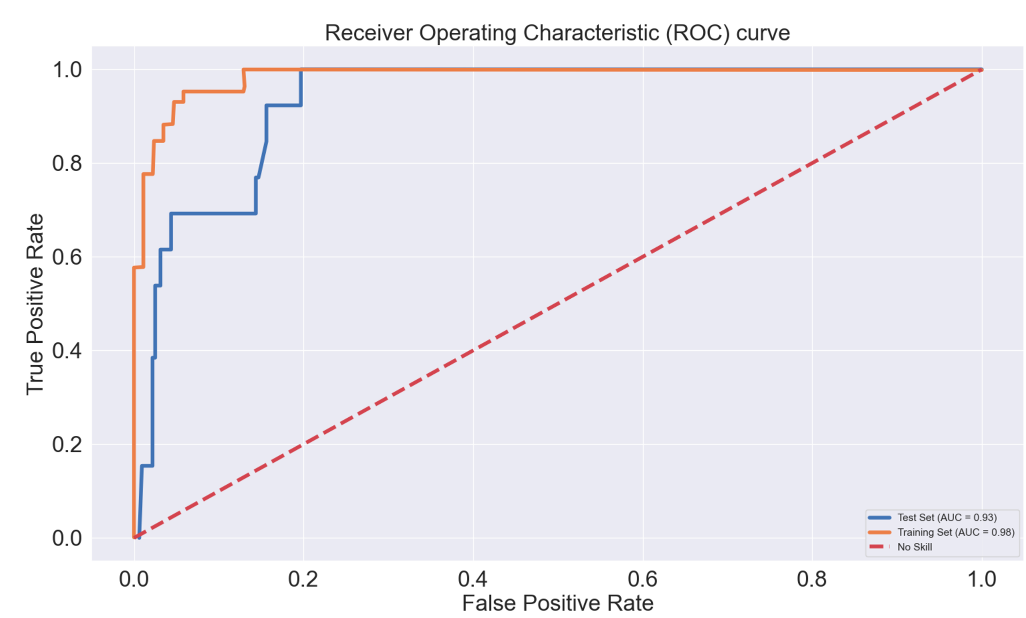
- In this study, we further enhance the models by incorporating lightning data as an additional feature. Lightning often precedes heavy rainfall, which can cause flash floods. We extracted relevant features from our dataset by creating 24-hour lighting vectors for each flood event, integrating the number of lightning strikes within a 10 km radius around each of the nine GNSS station at 1-hour time windows.
- We trained an SVM model using a standard k-fold cross-validation approach due to the limited amount of data available. As with limited data, the standard k-fold cross validation approach is a suitable choice as it provides good balance between computational cost and the ability to obtain meaningful results, while still allowing for an evaluation of the model's generalization performance.
- A Bayesian optimization was used to choose the optimal hyperparameters for our SVM model, which has been shown to be more efficient and effective than grid search in many cases, particularly for complex, high-dimensional models such as SVM.
- Our results showed that the model achieved impressive performance across multiple score metrics calculated from the imbalanced testing set (ratio of 56 quiet days to 1 flood event in the testing set), including an accuracy of 0.9913, F1 score of 0.7917, HSS score of 0.7875, precision of 0.6786, recall of 0.95, and TSS score of 0.9421.



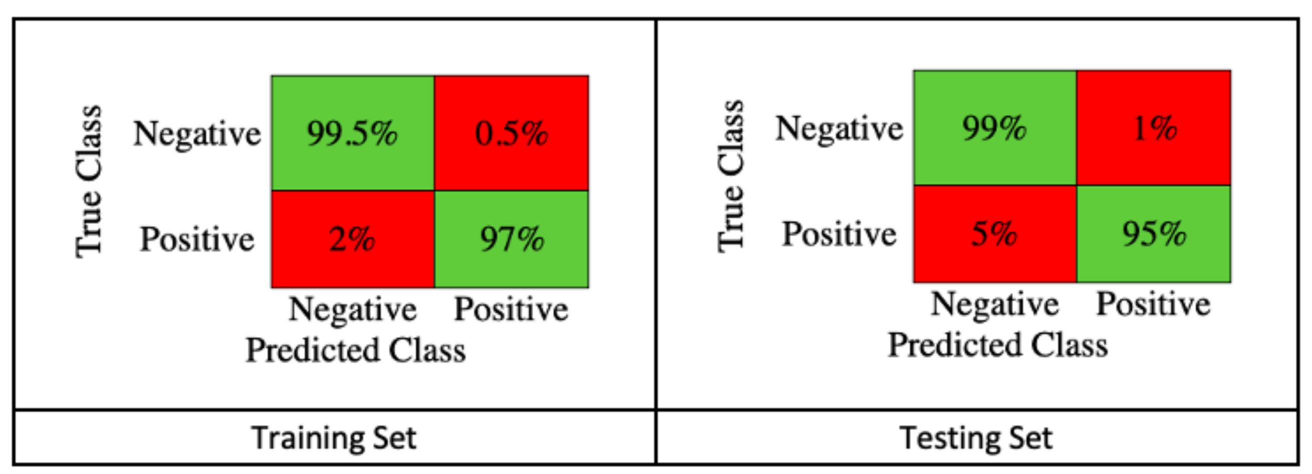
Imbalanced 5-Fold Cross Validation Results. The diagram illustrates the performance of the cross validation process with 5 subsets of the data, obtained through a randomized stratified sampling approach, allowing each iteration to randomly pick testing sets while still taking into account all the 9 stations (groups)

Comparison of skill score metrics for flash flood event prediction between the current SVM model and Ziv and Reuveni work. The results show an improvement in the accuracy of the current model in predicting flash flood events and non-floods, as indicated by the higher values in most skill scores





ROC model curve obtained during the hyperparameters optimization process



The confusion matrix for the SVM model results extracted from the training set (left), and the test set (right).