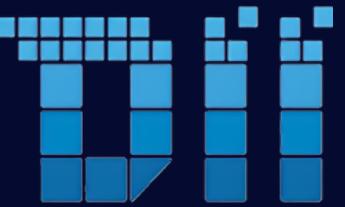




UNIVERSITÀ DI PISA



Avalanche Danger Level Prediction

FOR DRY-SNOW CONDITIONS

Author: Emanuele Respino
Data Mining and Machine Learning Course
A.Y. 2024/2025



INTRODUCTION



DATASET ANALYSIS



METHODOLOGY



INTERFACE



EXPLAINABILITY



RESULTS



INTRODUCTION



IN EUROPE ALONE, EAWS REPORTS AN AVERAGE OF **~100 AVALANCHE FATALITIES PER WINTER SEASON** DESPITE MODERN RESCUE AND SAFETY MEASURES



THE ALPINE REGION HOSTS **~375 MILLION OVERNIGHT STAYS ANNUALLY**: CRITICAL RELIABLE AVALANCHE FORECASTS ARE FOR BOTH LOCAL ECONOMIES AND VISITING SKIERS.



EACH DAY, EAWS FORECASTERS RELEASE A DANGER LEVEL TO **GUIDE PUBLIC AND COMMERCIAL DECISION-MAKING**.

INTRODUCTION



IN EUROPE ALONE, EAWS REPORTS AN AVERAGE OF **~100**
AVALANCHE FATALITIES PER WINTER SEASON DESPITE MODERN
RESCUE AND SAFETY MEASURES



TODAY'S BULLETINS DEPEND ON **EXPERT INTERPRETATION**
OF SNOWPACK OUTPUTS AND FIELD DATA, INTRODUCING
SUBJECTIVE VARIABILITY.



MANUAL WORKFLOWS CANNOT SCALE TO HOURLY
NOWCASTING AT INDIVIDUAL IMIS STATIONS, DELAYING
CRITICAL INFORMATION EXACTLY WHEN IT'S NEEDED MOST.

DATASET ANALYSIS

The Envidat dataset



TARGET:
**AVALANCHE DANGER
LEVEL RATING**

Categorical target variable (1–5) indicating overall snowpack stability, possible triggering events, expected slopes in which the avalanche can be generated. Only **certified** SLF avalanche danger levels.

METEOROLOGIC AL VARIABLES

Hourly weather measurements aggregated over 24 h to capture conditions driving snow processes.

SNOWPACK FEATURES

Layer-specific properties extracted from SNOWPACK's detailed snow-cover profiles.

AGGREGATED FEATURES

Aggregated versions of key variables (up to one week earlier).

DERIVED FEATURES

Higher-level summary values condensing complex physical processes.

CONTEXT FEATURES

Localization metadata and station attributes for tracking and grouping observations.

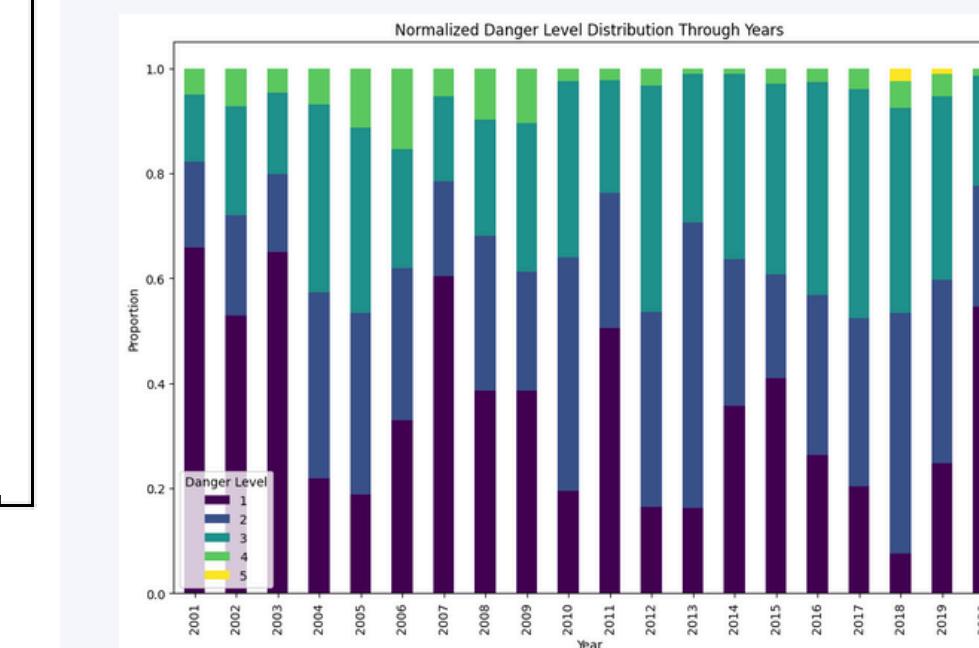
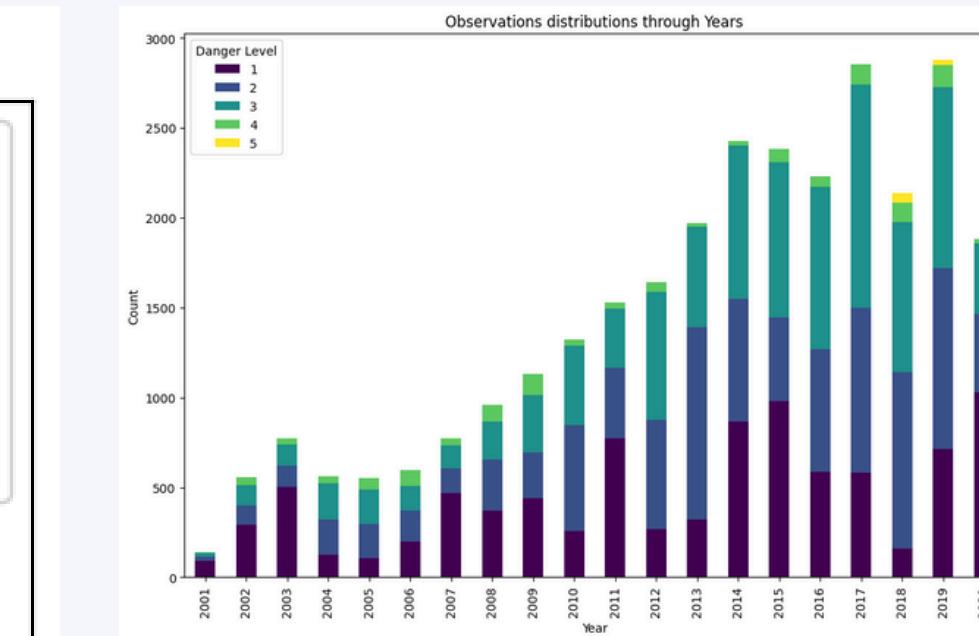
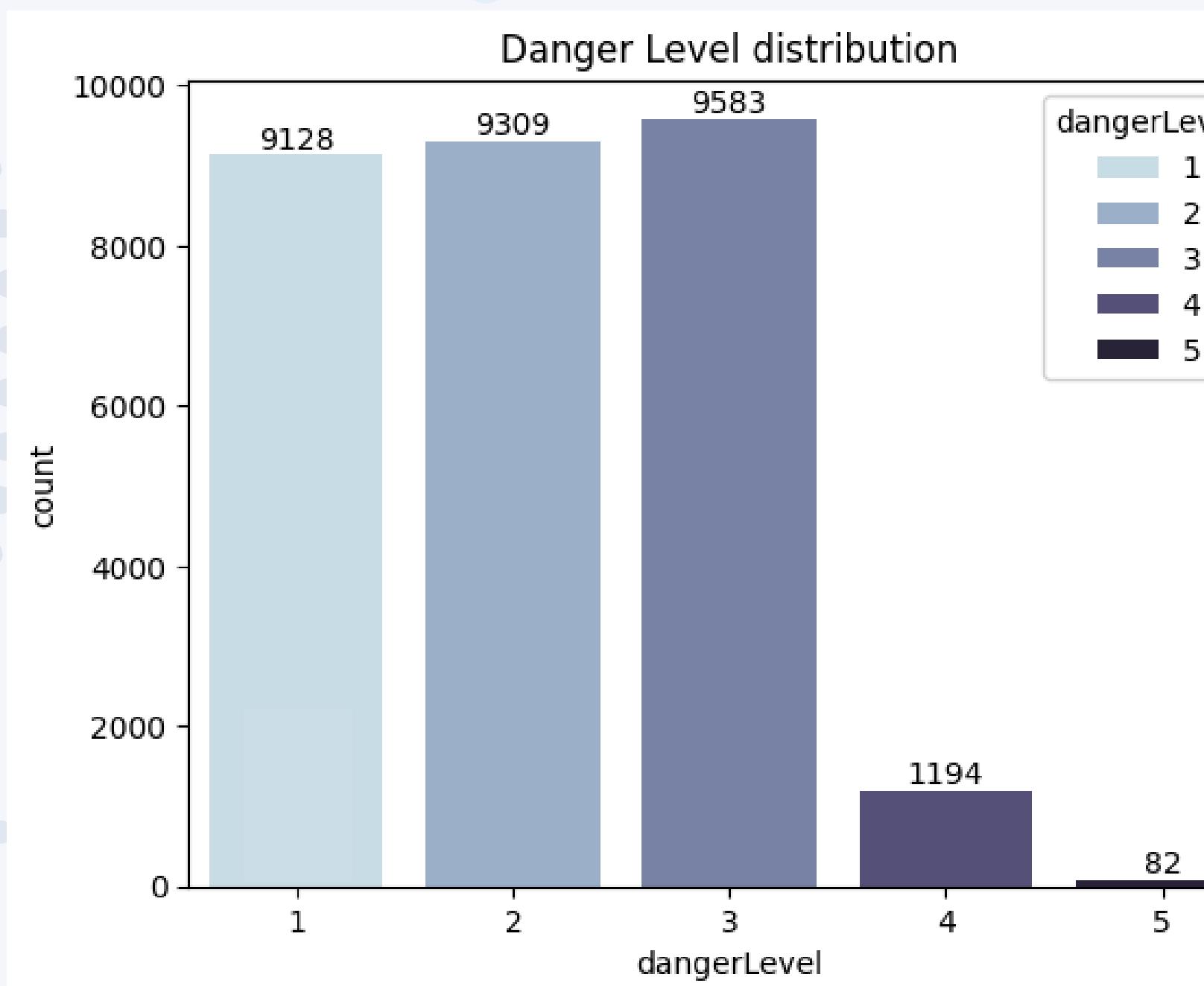
29 296 OBSERVATIONS | 79 FEATURES

WHERE: forecasted from different meteorological stations in Switzerland

WHEN: winter periods between 2001 and 2020

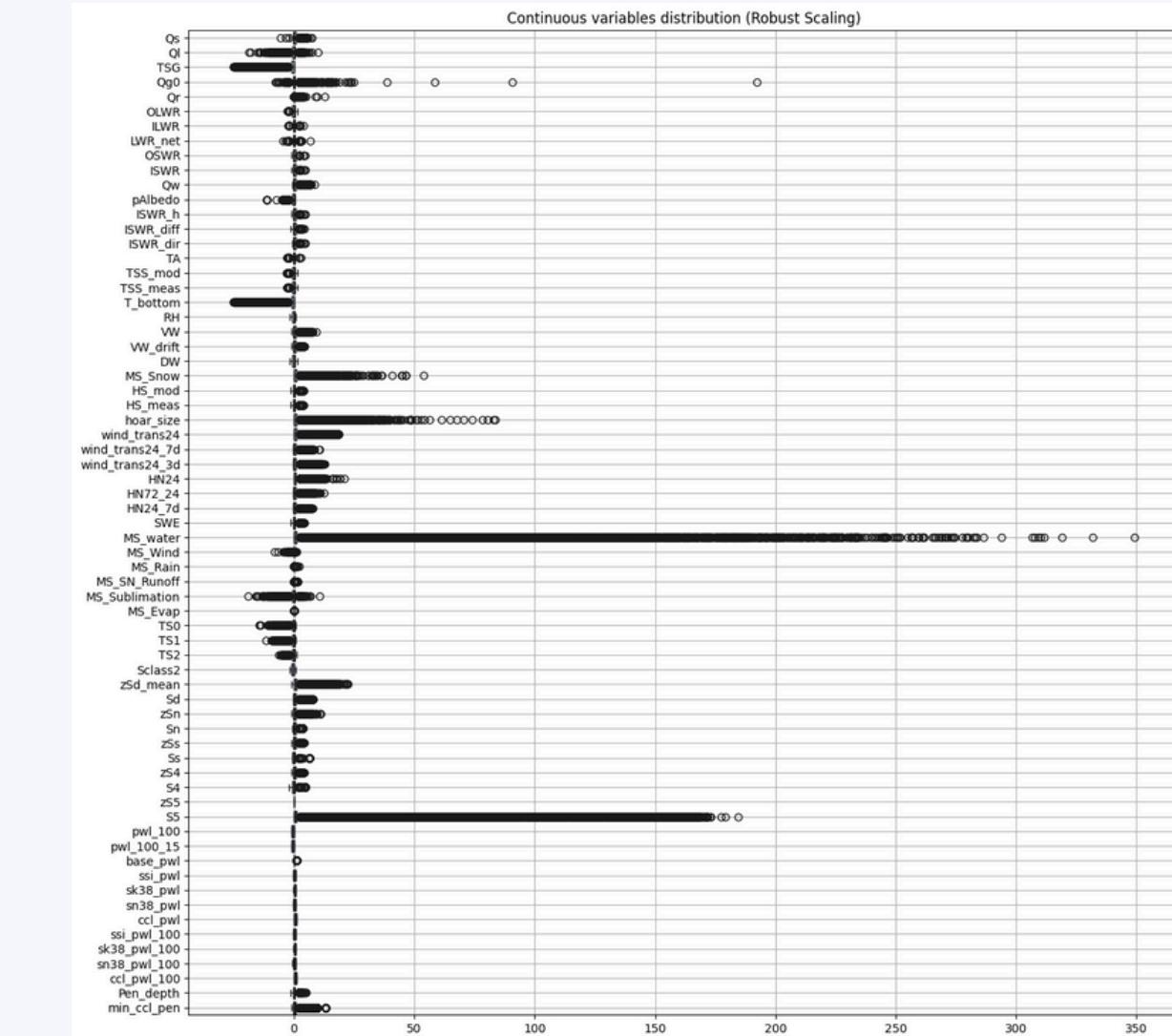
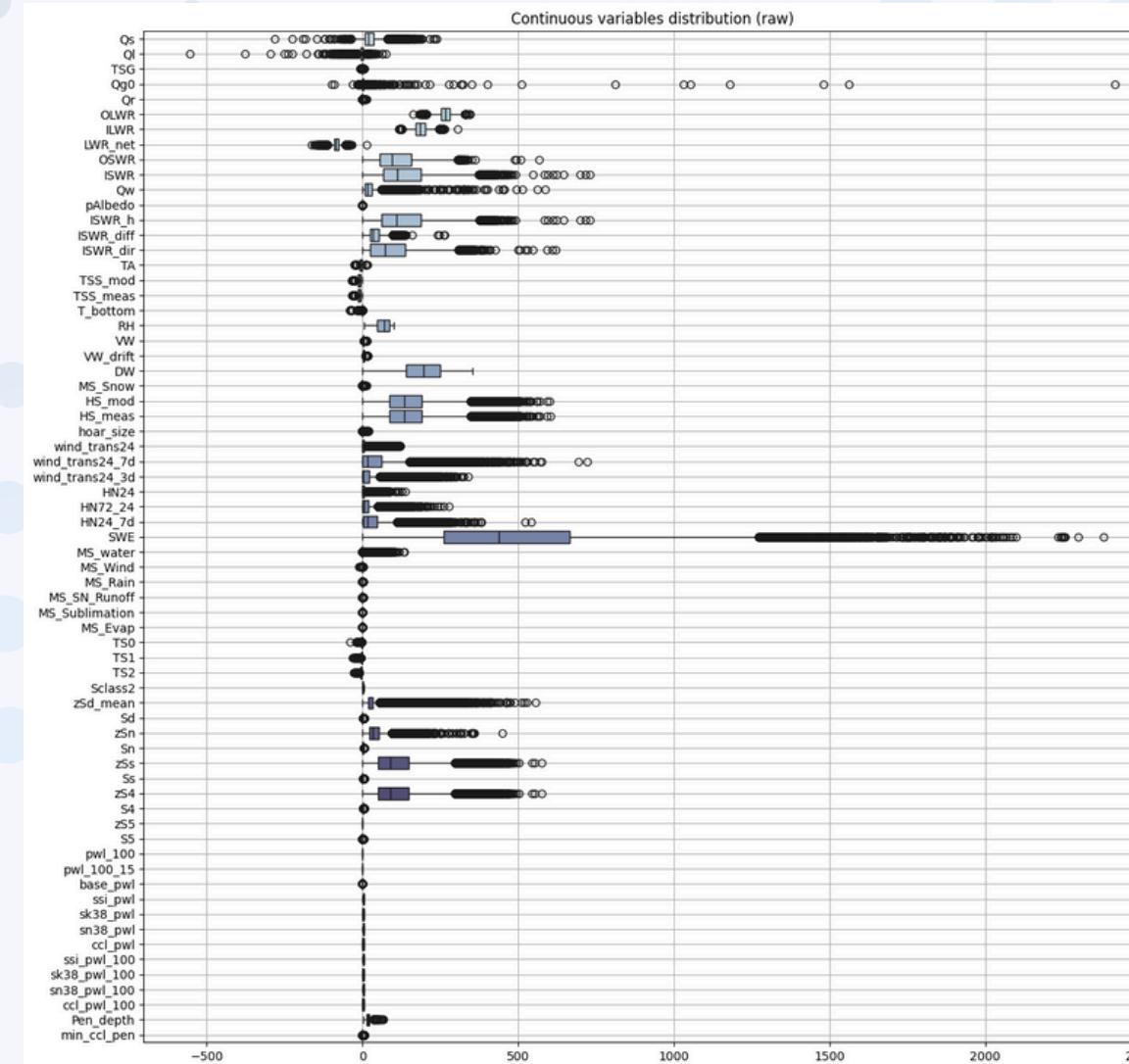
DATASET ANALYSIS

EDA - Target distribution



DATASET ANALYSIS

EDA - Data variety



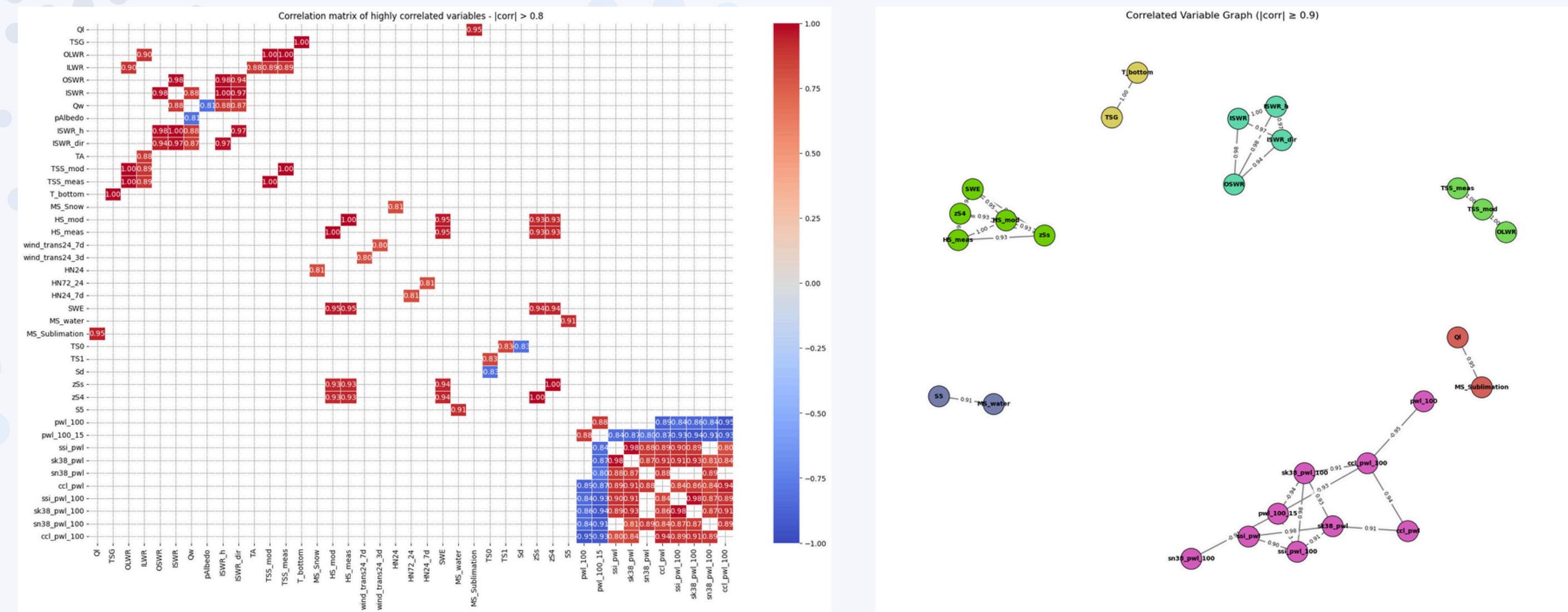
5 EMPTY
FEATURES

1 CONSTANT
FEATURE

77% OBSERVATIONS WITHOUT
MISSING VALUES

DATASET ANALYSIS

EDA - Features Correlation



7

GROUPS OF HIGHLY (>0.9)
CORRELATED FEATURES

~30

COMPONENTS NEEDED TO REACH
0.95 C.V. USING STANDARD SCALING

DATASET ANALYSIS

Preprocessing



29 296 OBSERVATIONS
79 FEATURES
1-5 CLASS TARGET

FEATURE PRUNING

- Removed empty and constant features
- dropped metadata to avoid bias
- kept datnum for chronological splits

MISSING-VALUES REMOVAL

Deleted all records with any missing values, without altering target-class balance.

CLASS MERGING

Combined danger levels 4 and 5 to increase sample size and reduce rare-class instability.

22 601 OBSERVATIONS
68 FEATURES
1-4 CLASS TARGET

DATASET ANALYSIS

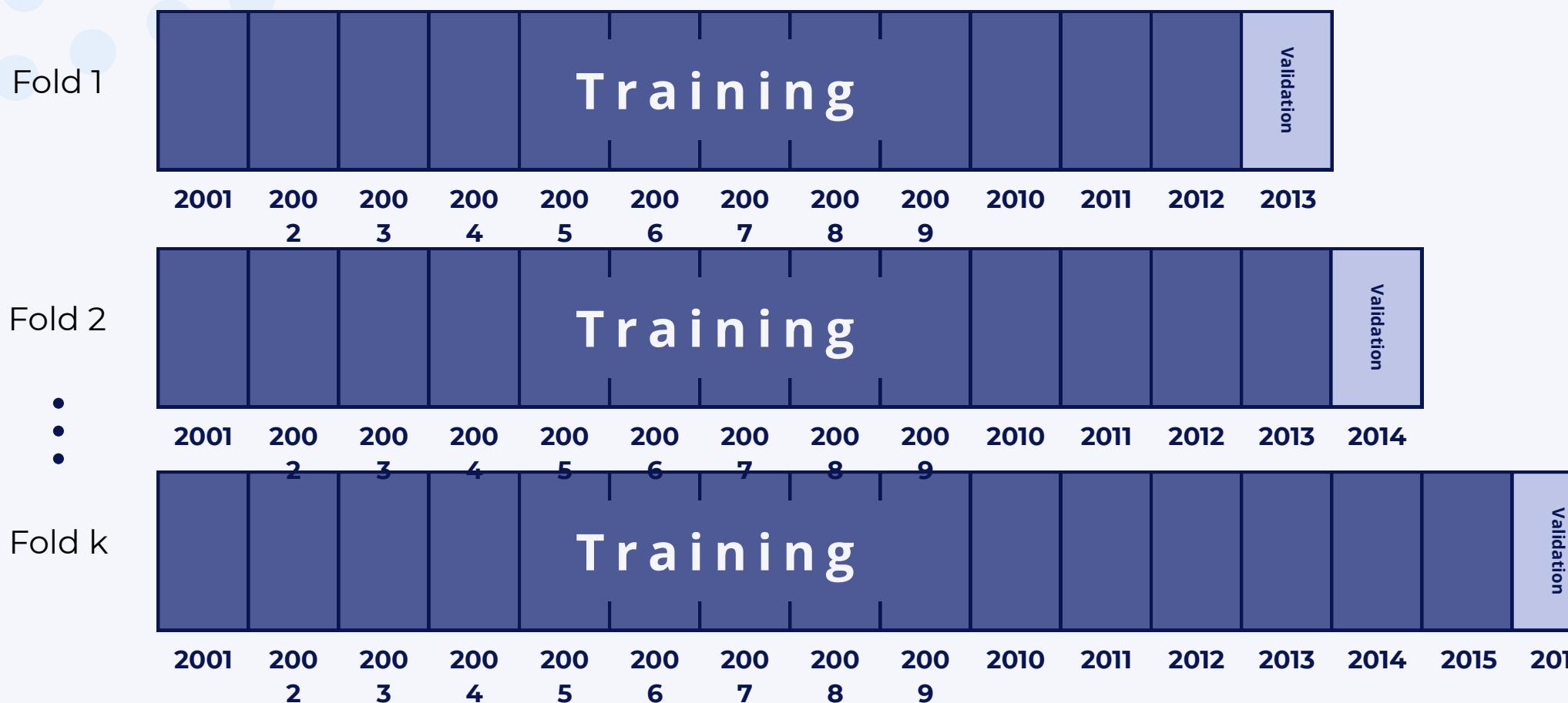
Splitting



Outer Hold-Out Split



Inner K-Fold CV



**GROUP OBSERVATIONS BY
WINTER SEASON (NOV–MAR)**

1

**SEASONS 2017–2019
OUT-OF-TIME TEST SET.**

2

**SEASONS 2001–2016
TRAIN ON SEASONS $\leq Y$ AND
VALIDATE ON SEASON $Y+1$.**

3

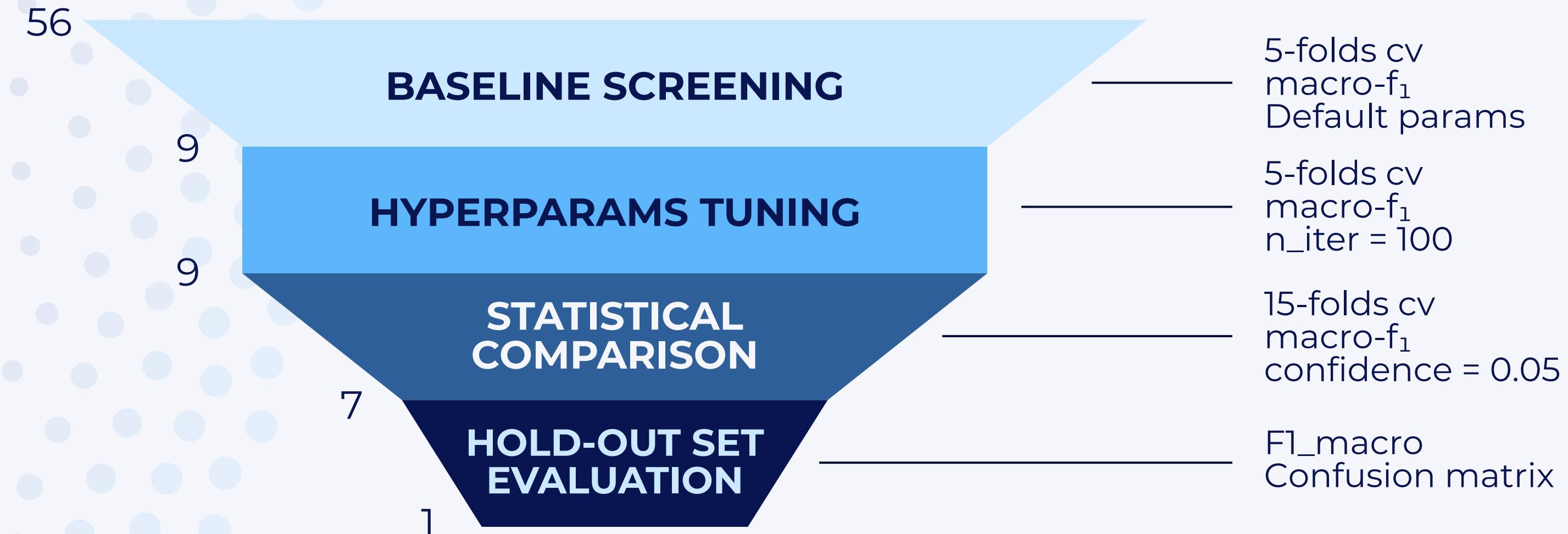
METHODOLOGY

Model Training



PIPELINES

1. **Sampling:** RandomOverSampler, SMOTE, None
2. **Scaling:** RobustScaler, StandardScaler, None
3. **Dimensionality red.:** PCA, Local-PCA, None
4. **Classifier:** RandomForest, XGBoost, SVM



METHODOLOGY

Model Optimization



HOLD-OUT SET EVALUATION

REMOVING
LESS-
IMPORTANT,
HIGHLY
CORRELATED
FEATURES

DETERMINING
MINIMAL
FEATURE
SUBSET SIZE

RFE

RE-TUNING

STATISTICAL
COMPARISON

FINAL
HOLD-OUT
EVALUATION

Drop low-rank
features with
 $|p| > 0.90$ to retain
only top
informative ones

Identify smallest
 k whose macro-
 F_1 matches full
set (Wilcoxon)

Use RFE to select
 k features and
update pipeline

Run randomized
search over
reduced pipeline

Compare 15-fold
CV macro- F_1
scores via
Wilcoxon test

Retrain on 2001–
2016 and test on
2017–2019,
reporting metrics

METHODOLOGY

Discriminating Class 5



Short **Second-stage** analysis on discriminating **Class 5**:

Margin Analysis

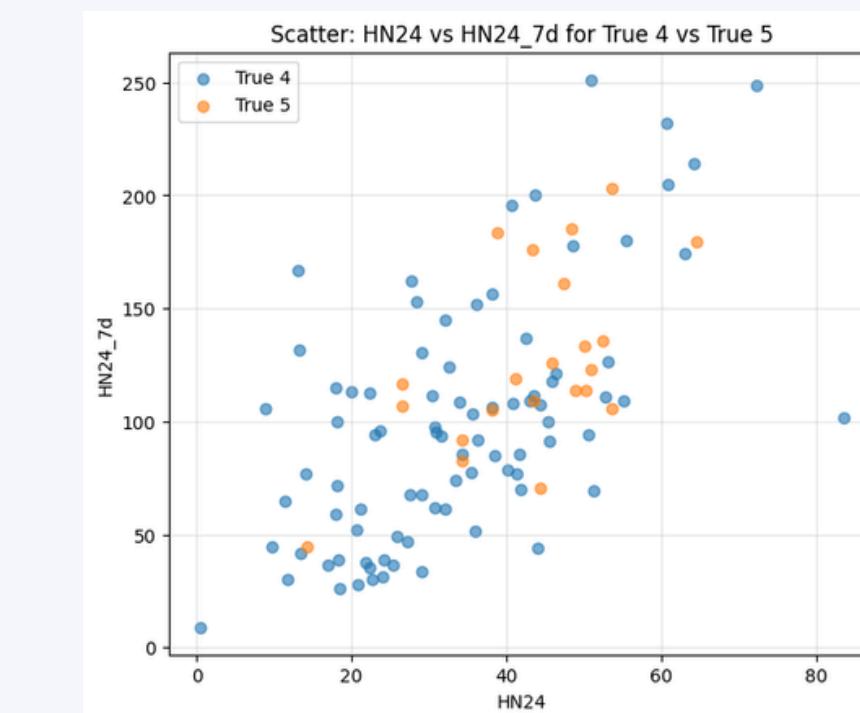
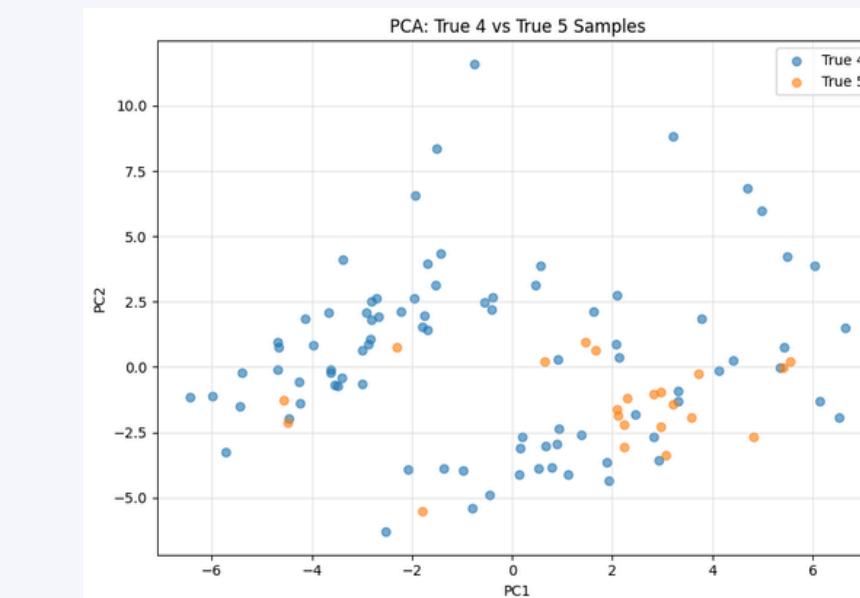
1. Compute $p_4 - p_3$ for true 4/5 samples
2. Rank by this margin
3. Find the threshold that maximizes F_1

Logistic Meta-Classifier

1. Logistic regression on $\{p_4, p_4 - p_3\}$ from true 4/5 training data
2. Map its binary output to 4 or 5 at test time

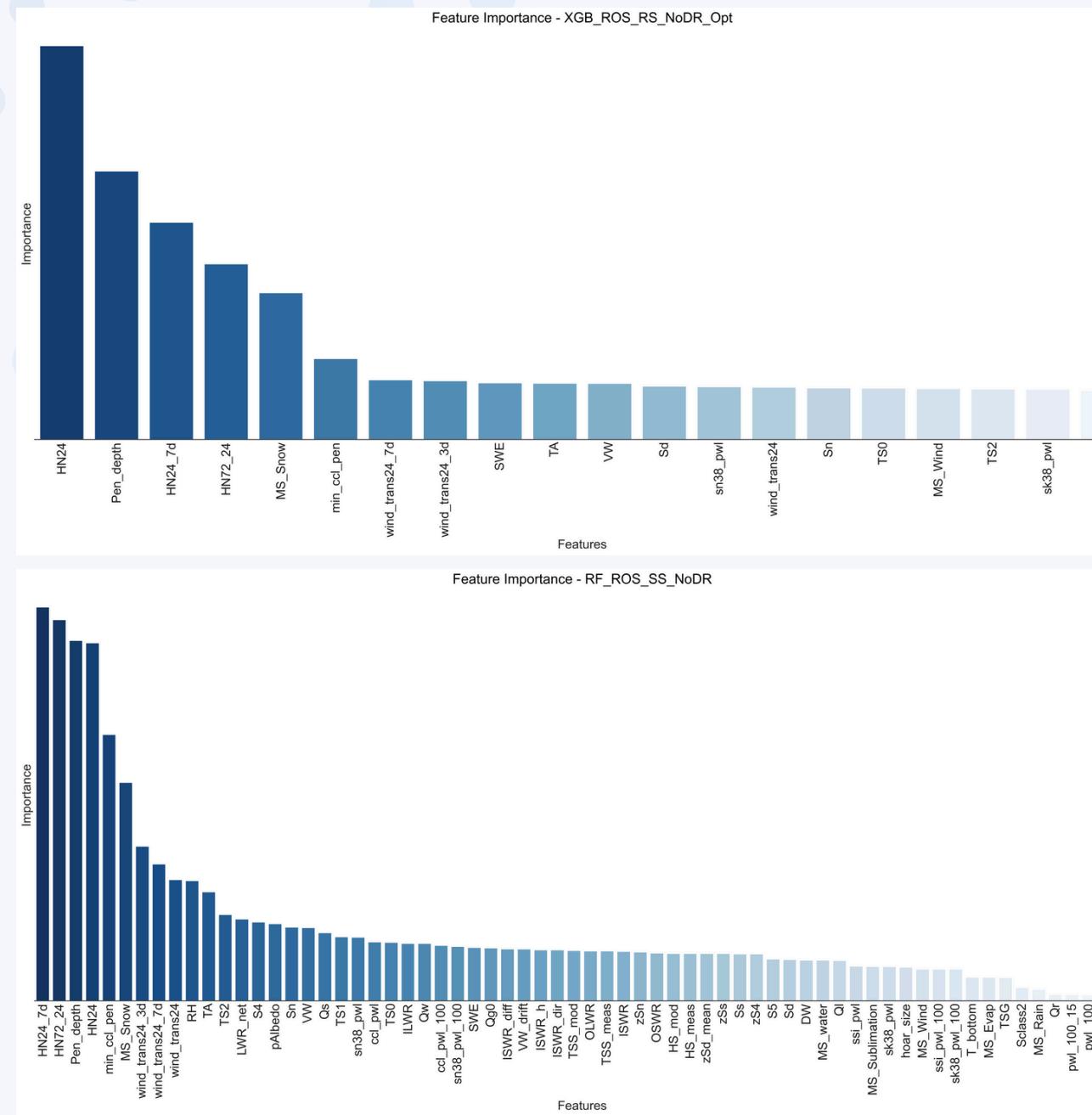
IsolationForest Outlier Detection

1. Fit an IsolationForest on true 4s (contamination = class 5 fraction)
2. Label test points with negative anomaly scores as 5

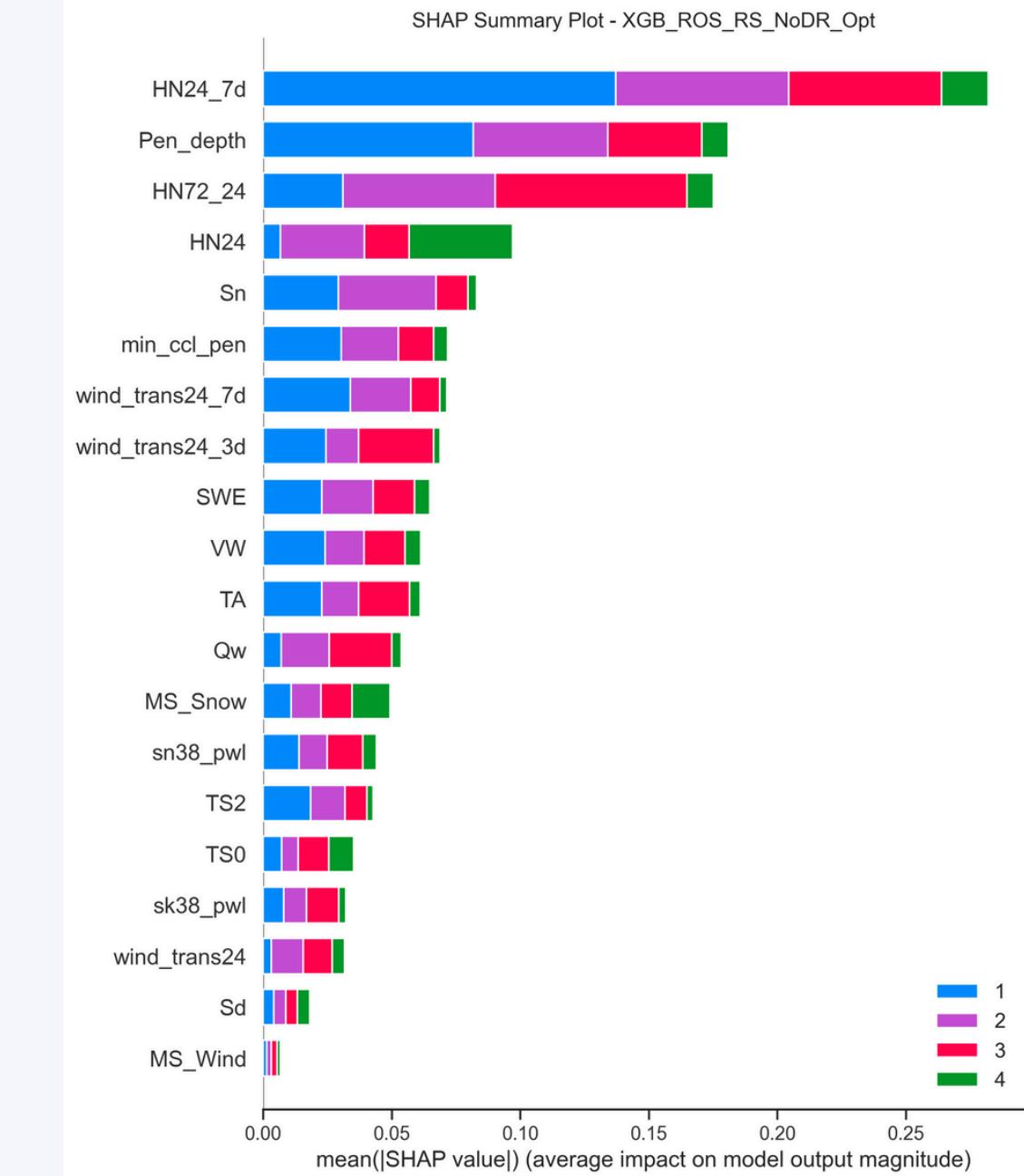


EXPLAINABILITY

Global



FEATURE IMPORTANCE



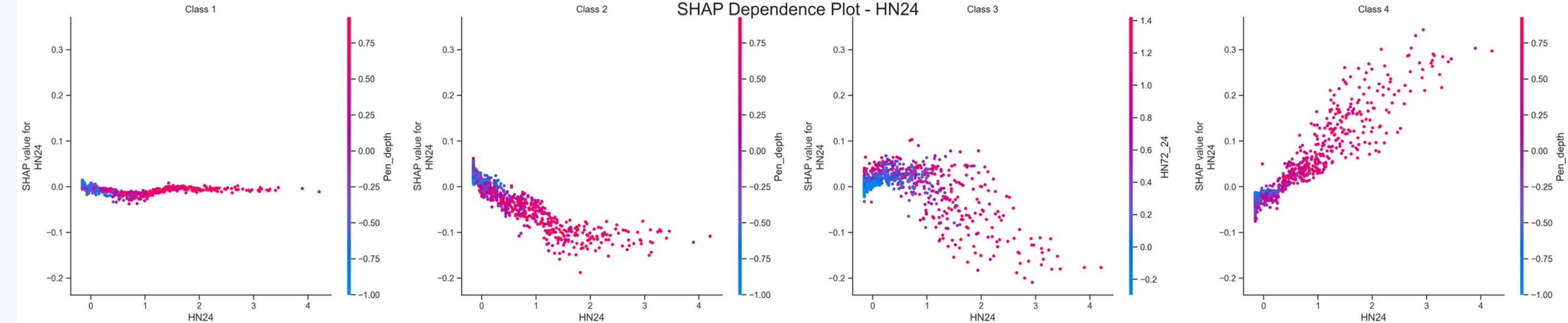
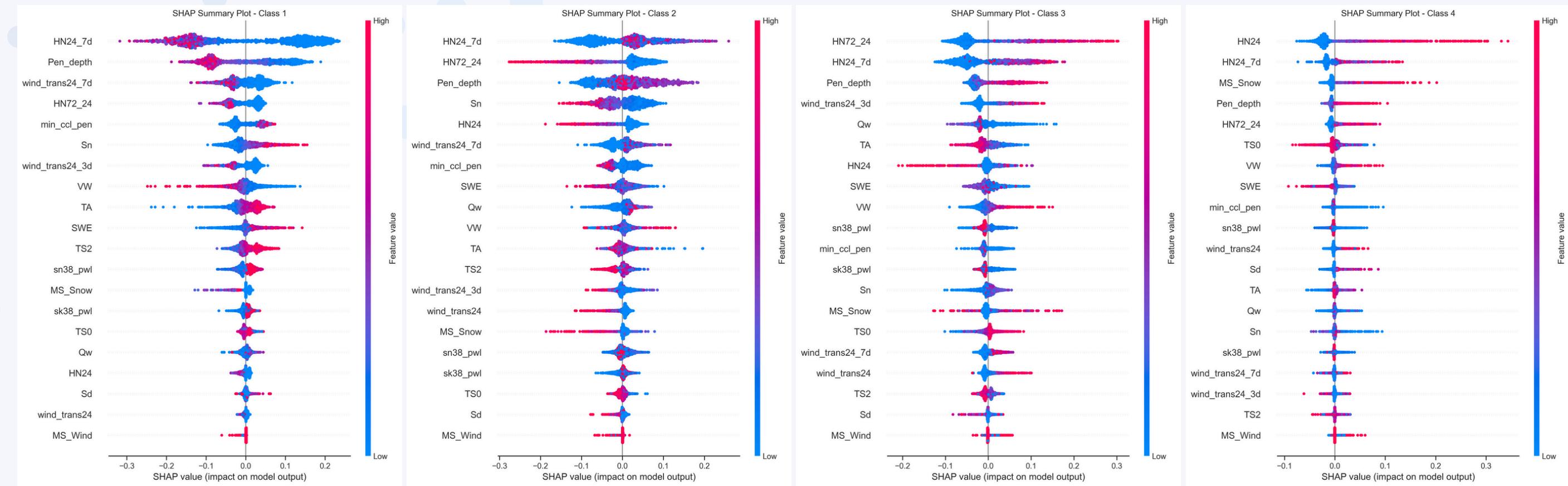
SHAP SUMMARY

EXPLAINABILITY

Global



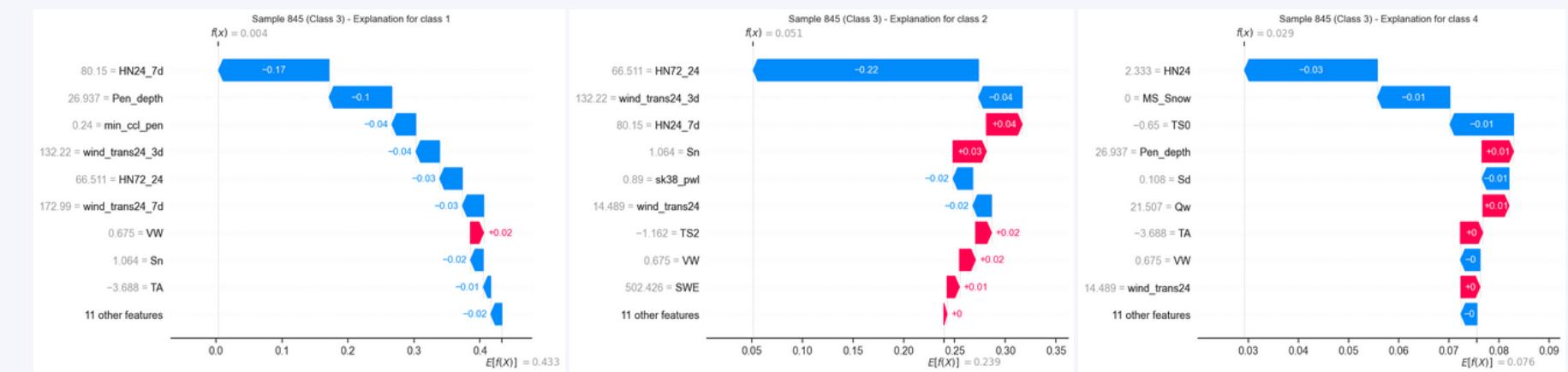
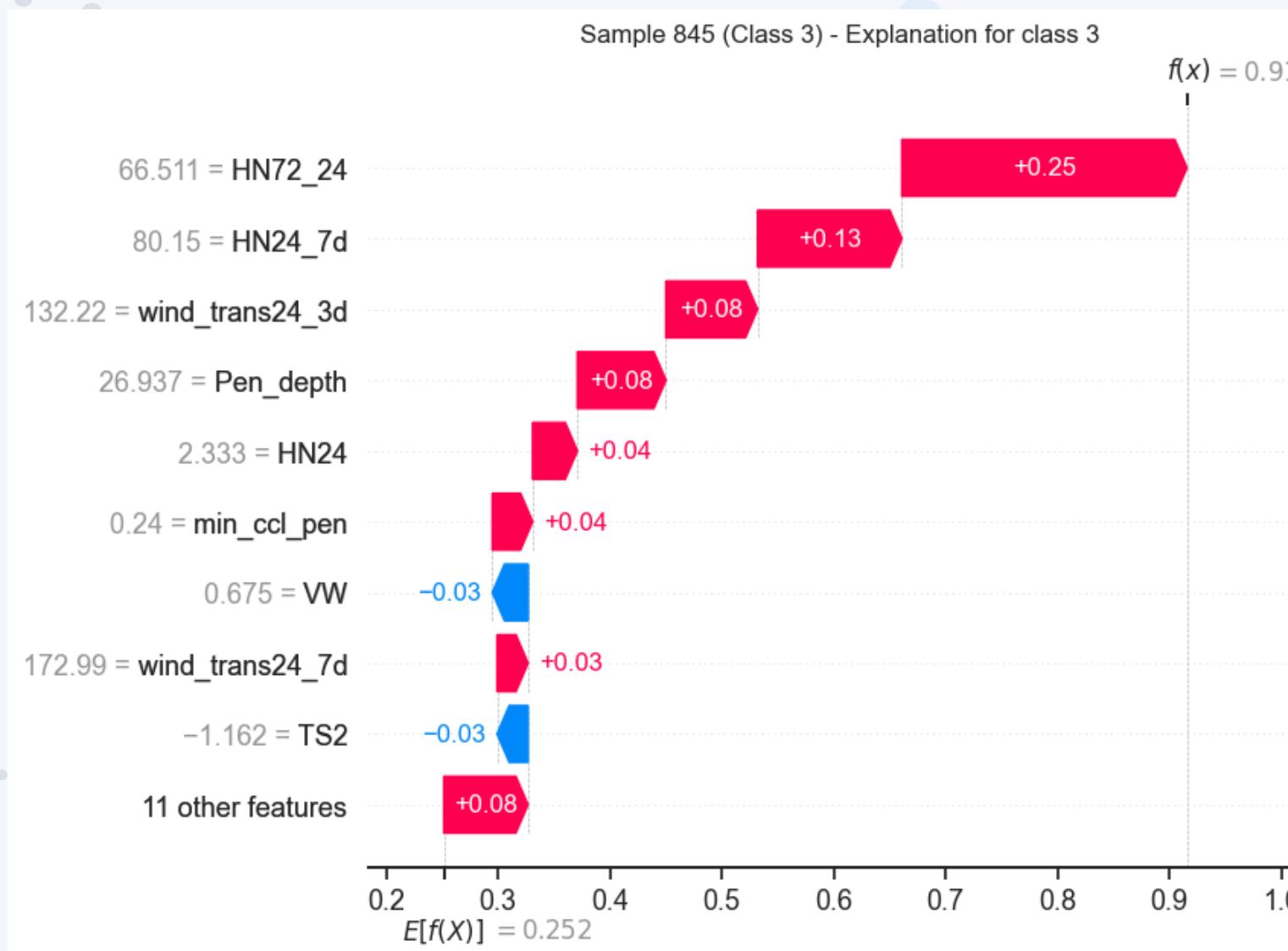
SHAP SUMMARY FOR EACH CLASS



DEPENDENCE PLOT

EXPLAINABILITY

Local



IMPORTANCE OF LOCAL EXPLAINABILITY

Break down individual predictions to show how feature values drive the model's decision for each observation.

PRACTICAL EXAMPLE (WATERFALL PLOT)

Class 3 sample to visualize each feature's positive or negative contribution to the prediction score across all classes.

RESULTS



IMPACT OF PIPELINE DESIGN CHOICES

Model choice: Decision-tree methods outperforms others by capturing non-linear interactions and tolerating redundant features.

Oversampling: RandomOverSampler consistently beat SMOTE, likely because SMOTE creates unrealistic synthetic meteorological samples.

Scaling: No clear pattern, since tree-based models (RF, XGB) performances were not influenced by this step.

Dimensionality reduction: Pipelines without DR often performed best; group-wise PCA outperforms global PCA when used.

Pipeline	Accuracy	F ₁ -macro	Balanced Acc.	MCC
RF_ROS_RS_NoDR	0.754209	0.710624	0.698313	0.641582
RF_ROS_SS_NoDR	0.754563	0.710500	0.697977	0.642067
RF_SMOTE_NoS_L-PCA	0.742867	0.686936	0.682776	0.626233
XGB_NoR_RS_NoDR	0.739677	0.661388	0.645076	0.618530
XGB_NoR_NoS_L-PCA	0.745880	0.657204	0.641983	0.627126
XGB_ROS_RS_NoDR	0.742158	0.700881	0.706312	0.627547
Paper_Optimized	<i>0.748361</i>	<i>0.722989</i>	<i>0.731236</i>	<i>0.636519</i>



PERFORMANCES CONSIDERATIONS

XGB_ROS_RS_NoDR: achieved >50 % recall on class 4 in hold-out. ROC-AUC=0.94 but threshold sweeps (F_1 -, precision-, recall-optimized) failed to yield an unbiased production rule

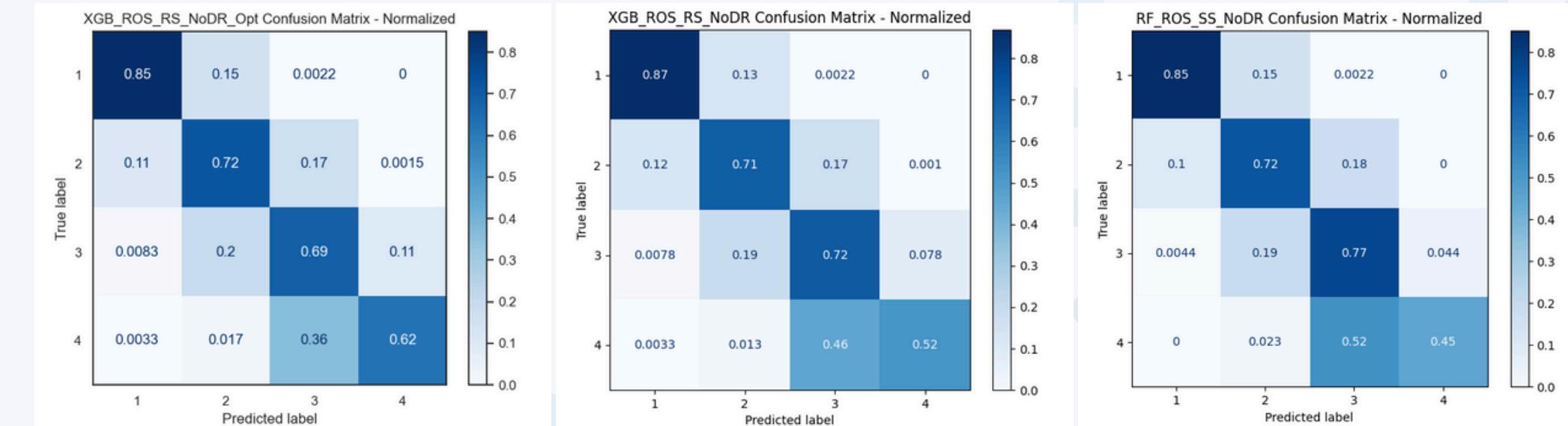
Model optimization:

- ↑ Balanced accuracy, recall on class 4
- ↓ Overall accuracy, F_1 -macro, MCC

Accepting minor overall performance drop may be worthwhile to **better detect high-risk days.**

Class 5 discrimination:

True-5 samples are too similar to true-4 in feature space; probability- or outlier-based methods cannot reliably separate them.



Pipeline	Accuracy	F_1 -macro	Balanced Acc.	MCC
XGB_ROS_RS_NoDR_Opt	0.734361	0.700612	0.720076	0.618952
XGB_ROS_RS_NoDR	0.742158	0.700881	0.706312	0.627547
RF_ROS_SS_NoDR	0.754563	0.710500	0.697977	0.642067

INTERFACE



IceSentinel
An Avalanche Danger Level Classifier

Select model

- Paper_Optimized
- RF_ROS_SS_NoDR
- XGB_ROS_RS_NoDR
- XGB_ROS_RS_NoDR_Opt

Deploy :

LOAD DATA

Upload CSV data

Drag and drop file here
Limit 200MB per file + CSV

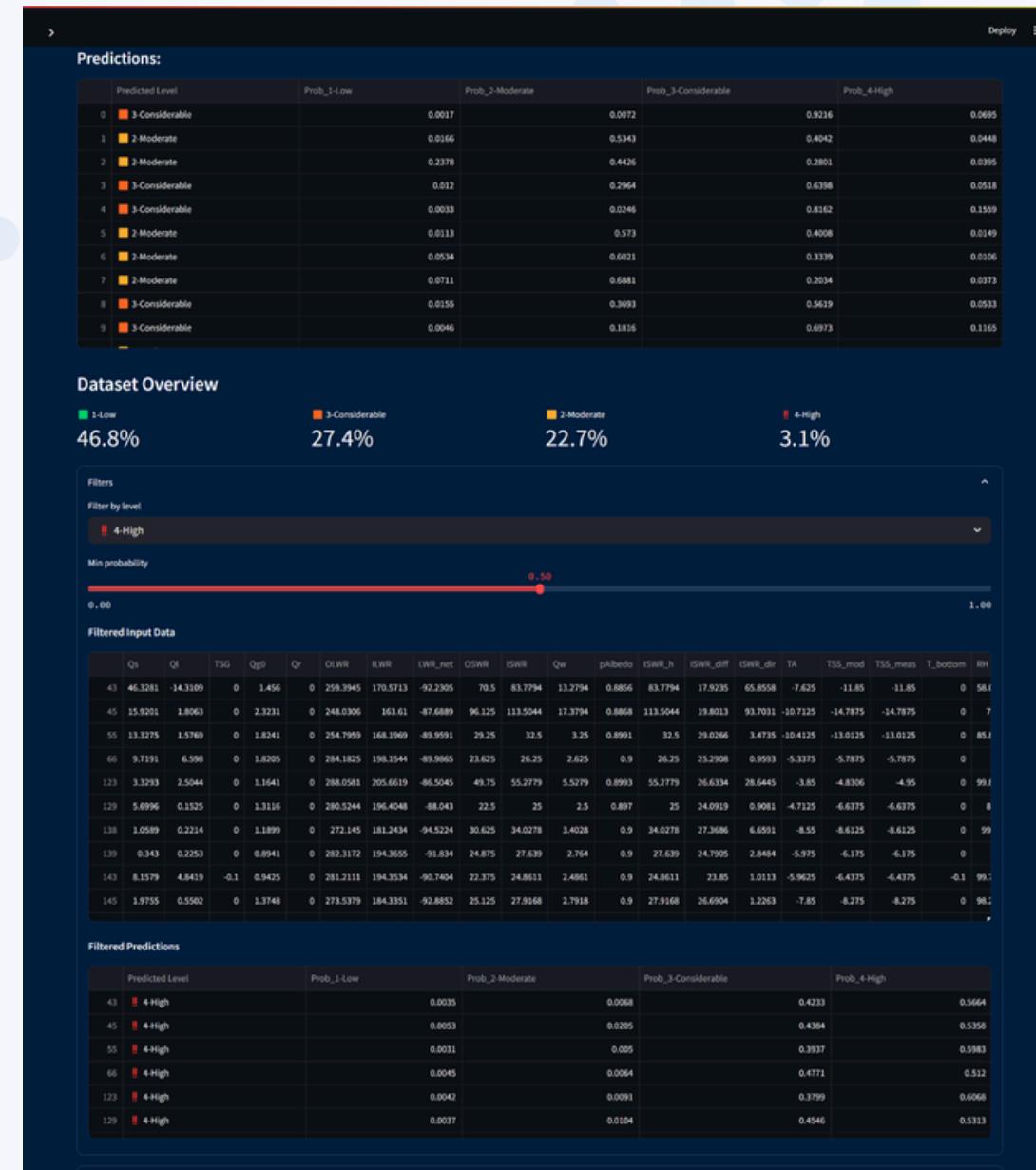
Browse files

SELECT MODEL

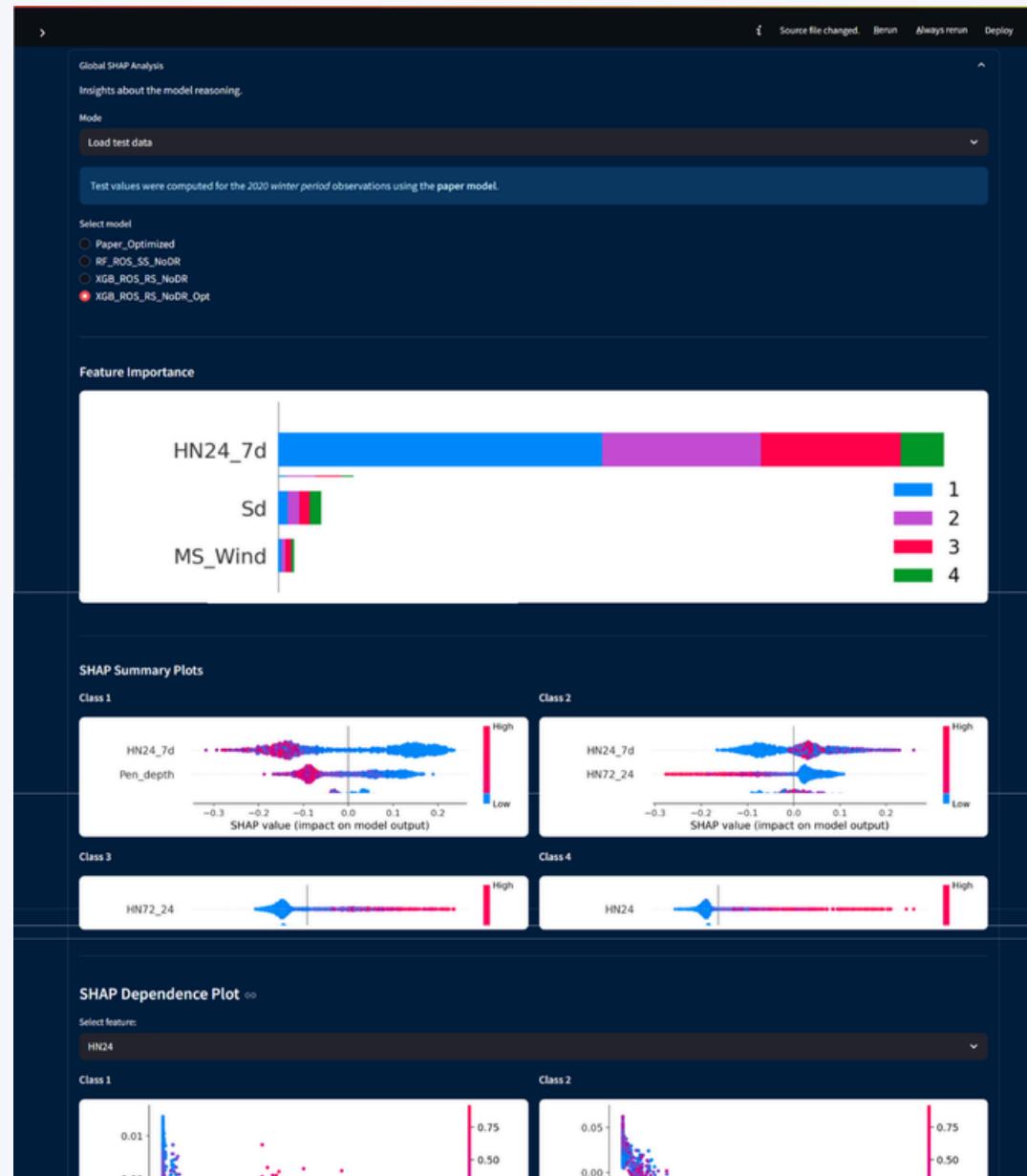
Input data:

	Qs	QI	TSG	Qg0	Qr	OLWR	ILWR	LWR_net	OSWR	ISWR	Qw	pAlbedo	ISWR_h	ISWR_diff	ISWR_dir	TA	T
0	18.6441	-4.3056	0	0.8961	0	271.7681	185.2941	-90.1754	29.25	32.5	3.25	0.8866	32.5	29.8431	2.657	-4.1125	
1	24.4668	-0.8685	-0.2	1.4036	0	258.6682	183.2321	-79.0965	45.25	50.2779	5.0279	0.8785	50.2779	34.8964	15.3815	-4.7875	
2	75.0902	-10.7519	-0.875	13.9625	0	267.421	182.4045	-88.66	43	47.7986	4.7986	0.9	47.7986	34.9855	12.8131	-6.2125	
3	47.3756	1.792	0	1.4603	0	253.1894	183.8658	-72.9964	41.25	45.9613	4.7113	0.8776	45.9613	34.9478	11.0134	-4.6125	
4	22.1523	4.5406	0	1.4706	0	269.8033	189.2884	-84.2963	152.375	169.3055	16.9305	0.9	169.3055	24.888	144.4178	-6.525	
5	22.227	-0.8749	-0.175	1.5779	0	244.6698	168.8755	-79.168	76.625	93.9168	17.2918	0.8646	93.9168	22.2191	71.6975	-6.3125	
6	51.415	-2.03	0	1.614	0	254.0714	175.0023	-82.5651	60	69.888	9.888	0.8904	69.888	16.355	53.533	-5.8125	
7	16.1195	1.9128	-0.2	1.3875	0.7923	235.6425	169.9211	-69.116	95.375	118.4733	23.0983	0.852	118.4733	17.4156	101.0576	-6.3375	
8	29.5555	5.9209	0	1.5045	0	229.5653	169.1619	-63.7828	92.625	115.1998	22.5748	0.8535	115.1998	18.7769	96.4229	-7.05	
9	27.8986	4.401	0	1.3745	0	233.4845	169.6995	-67.1752	81.5	99.8859	18.3859	0.864	99.8859	17.2843	82.6018	-7.5625	

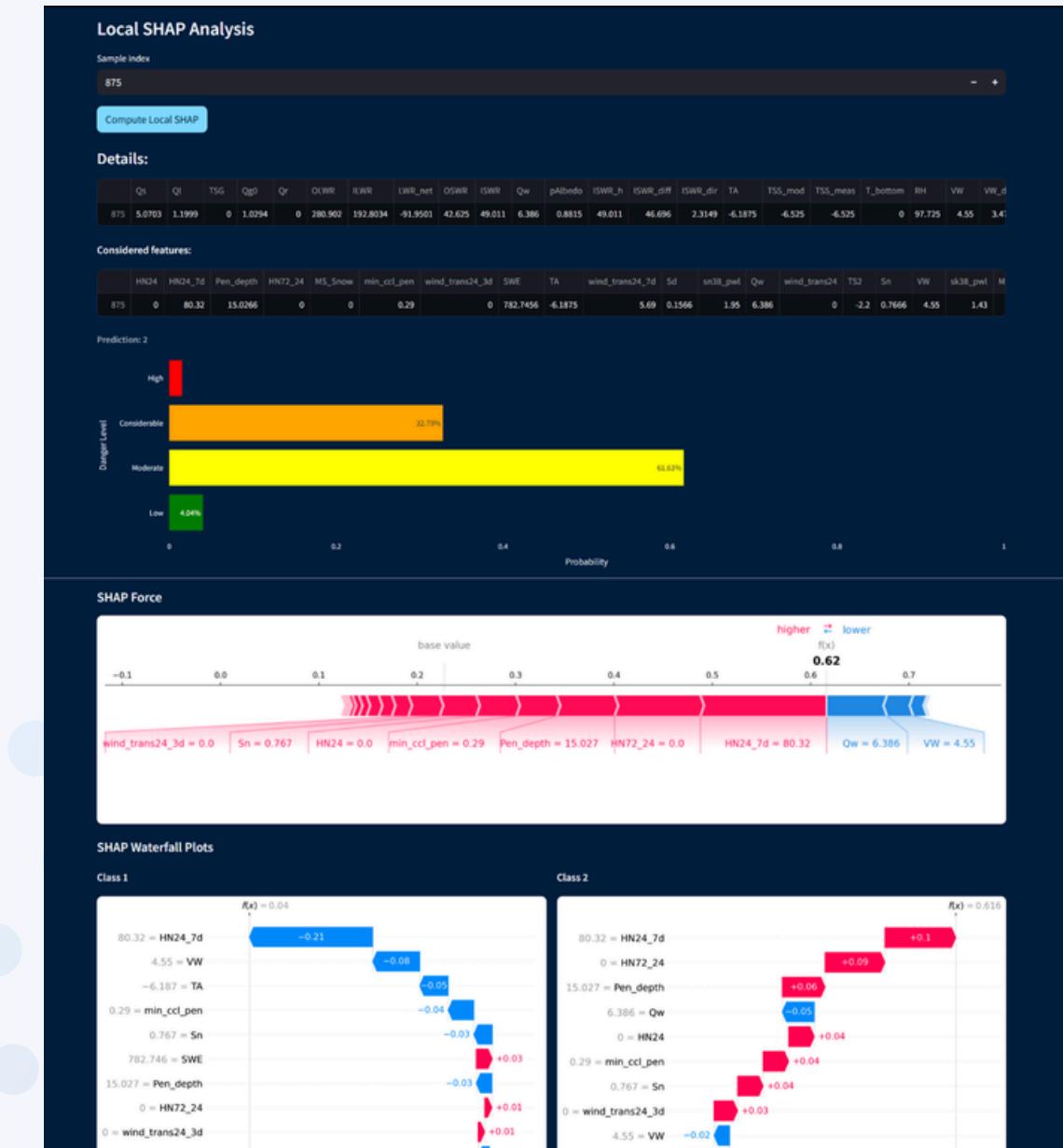
INTERFACE



PREDICTIONS



GLOBAL EXPLANATIONS



LOCAL EXPLANATIONS

Main References

Pérez-Guillén, C., Techel, F., Hendrick, M., Volpi, M., van Herwijken, A., Olevski, T., Obozinski, G., Pérez-Cruz, F., and Schweizer, J.: **Data-driven automated predictions of the avalanche danger level for dry-snow conditions in Switzerland**, Nat. Hazards Earth Syst. Sci., 22, 2031–2056, 2022.
<https://doi.org/10.5194/nhess-22-2031-2022>

Pérez-Guillén, C., Techel, F., Volpi, M., and van Herwijken, A.: **Assessing the performance and explainability of an avalanche danger forecast model**, Nat. Hazards Earth Syst. Sci., 25, 1331–1351, 2025.
<https://doi.org/10.5194/nhess-25-1331-2025>

Pérez-Guillén, C., Techel, F., Hendrick, M., Volpi, M., van Herwijken, A., Olevski, T., Obozinski, G., Pérez-Cruz, F., Schweizer, J. (2022). **Weather, snowpack and danger ratings data for automated avalanche danger level predictions**. EnviDat.
<https://www.doi.org/10.16904/envidat.330>



THANK YOU