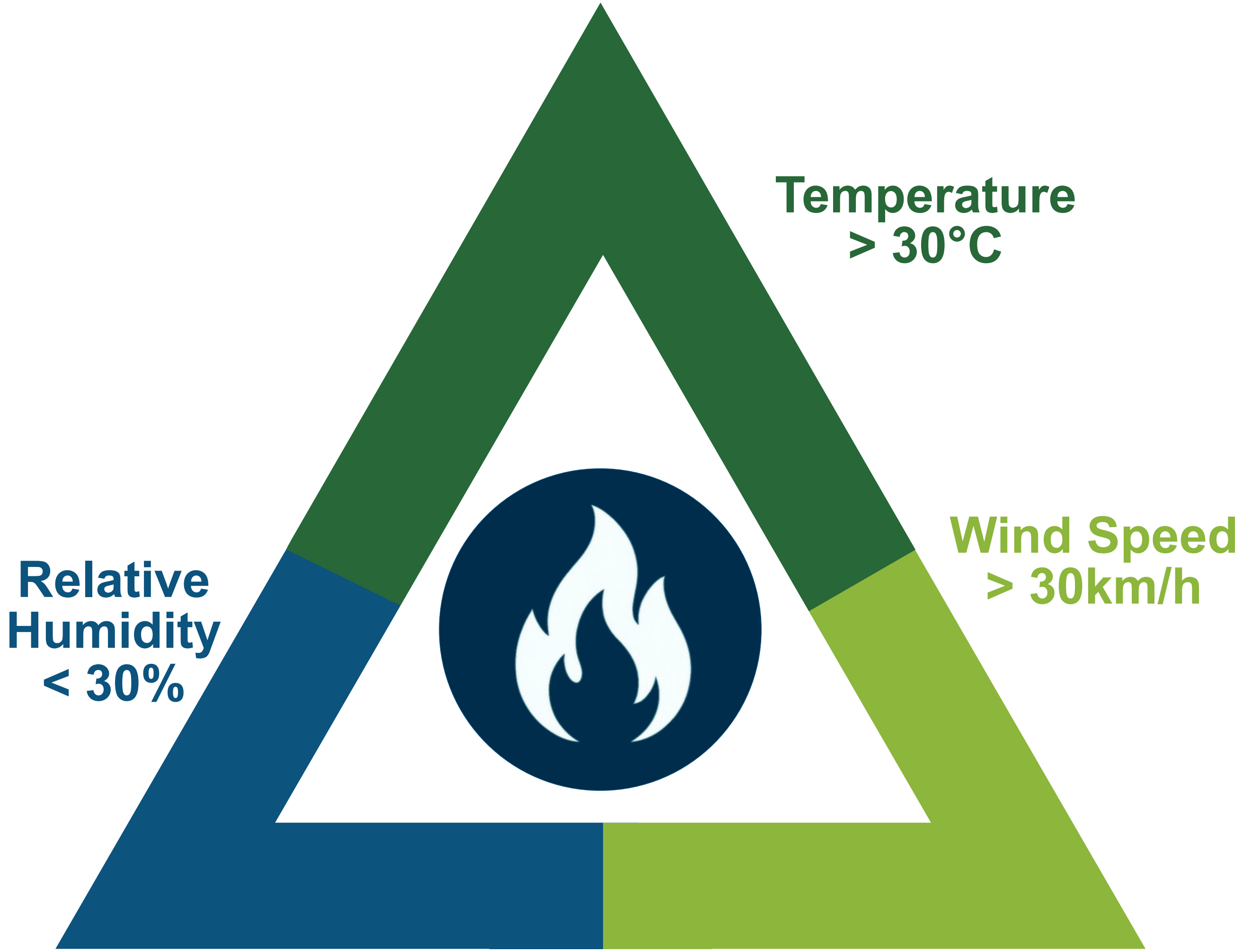
Designing and Explaining Temporal Deep Learning Models for Wildfire Danger Prediction

Pauline Becker
Bachelor Thesis, 07th August 2025

Wildfires in the Mediterranean Region

Megafire Triangle: 30-30-30 Rule

- Fire-prone Region
- Mediterranean has seen a sharp rise in burned area
  - strongly linked to anthropogenic climate change



Farid et.al., 2024

Drivers of Fire

Fuel

- Combustible vegetation (grasses, shrubs, trees) determines fire potential
- Act as binary constraints: if non-burnable areas dominate, no fire occurs regardless of weather

Meteorology

- High temperatures and low humidity dry out fuels
- Key driver in determing wheter an ignition occurs or fire can spread

Human Factors

- Major ignition source in fire-prone regions, e.g. Mediterranean

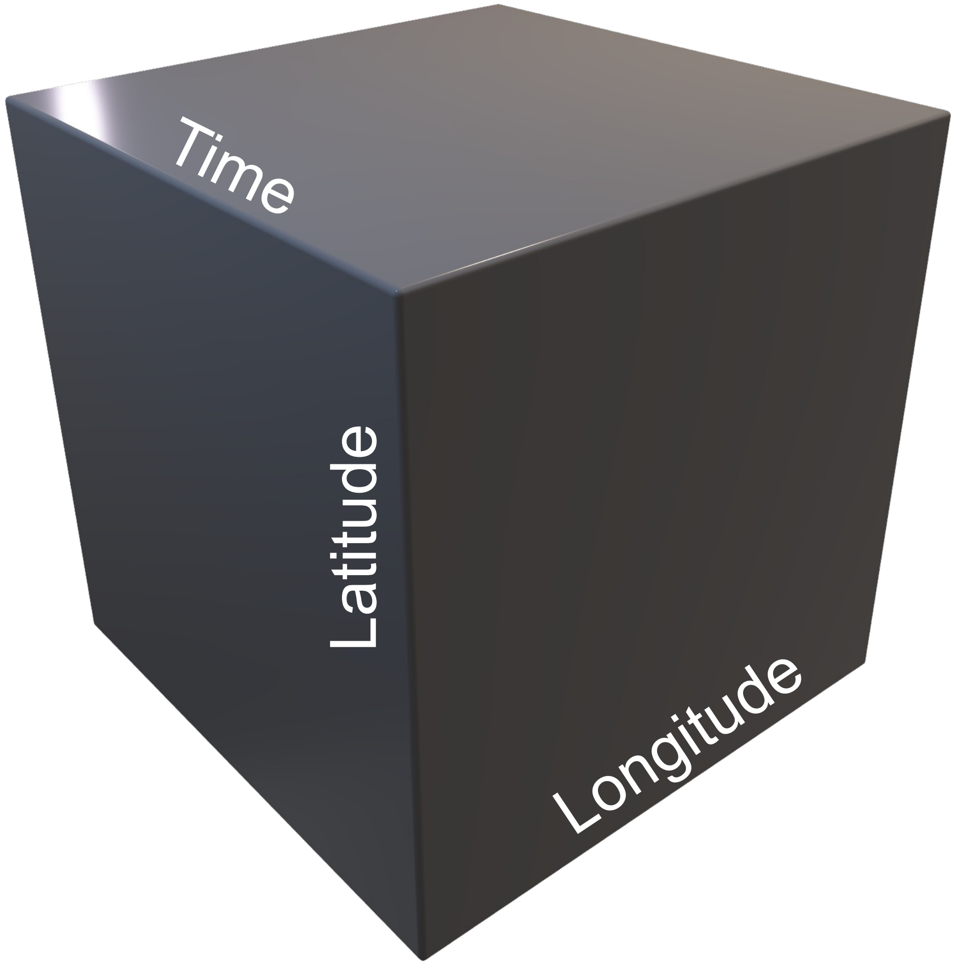
Topographical Factors

- Steeper slopes promote faster uphill fire spread
- Elevation influences fuel types and local weather conditions
- Static nature results in lower explanatory power

Mesogeos Dataset

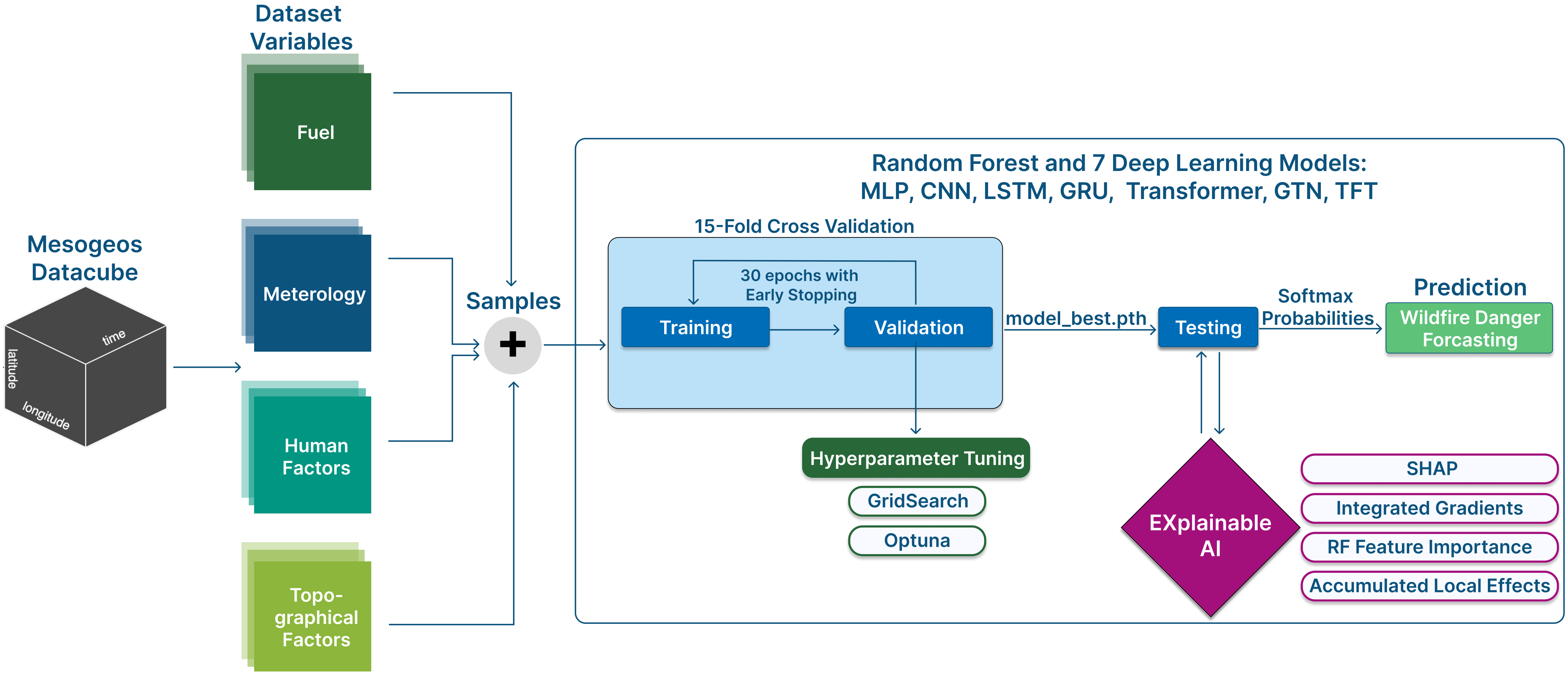
Spatio-temporal datacube covering Mediterranean region from 2006-2022

- Daily temporal resolution
- 1 km × 1 km spatial resolution
- Includes 24 variables: meteorology, vegetation, land cover, human activity
- Over 47.8 billion data points per dynamic variable
- Missing values (mainly from satellite gaps, e.g., cloud cover) were filled using the temporal mean per feature



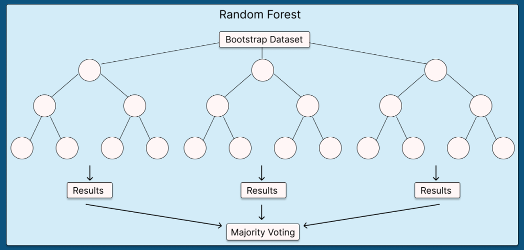
Kondylatos, Prapas, Camps-Valls, & Papoutsis, 2023

Machine Learning Pipeline for Wildfire Danger Prediction

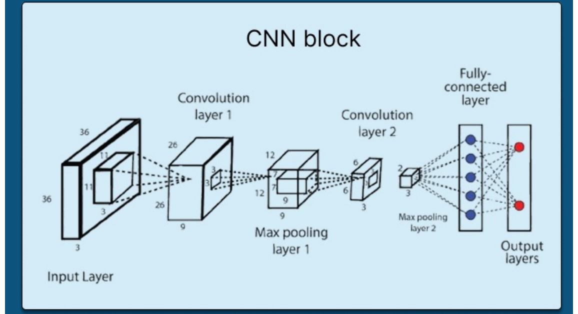


Machine Learning Models for Wildfire Danger Prediction

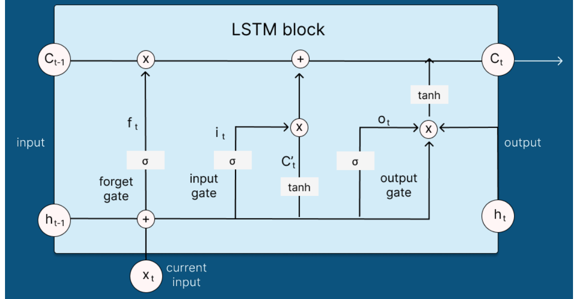
Random Forest (RF)



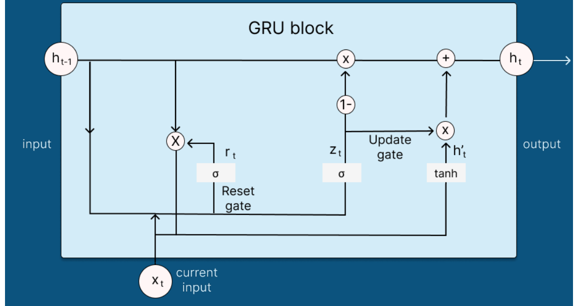
Convolutional Neural Network (CNN)



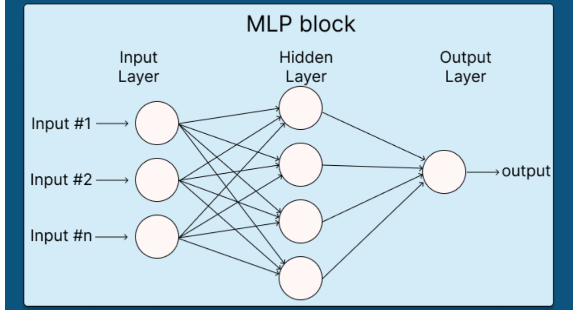
Long Short-Term Memory (LSTM)



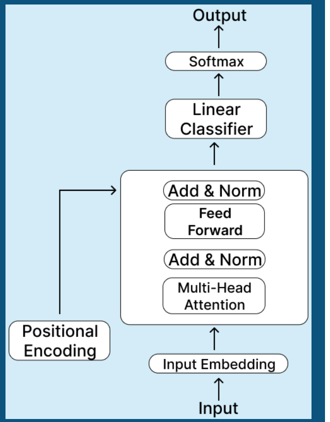
Gated Recurrent Unit (GRU)



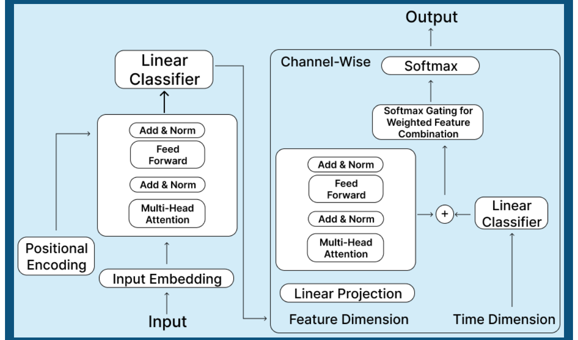
Multilayer Perceptron



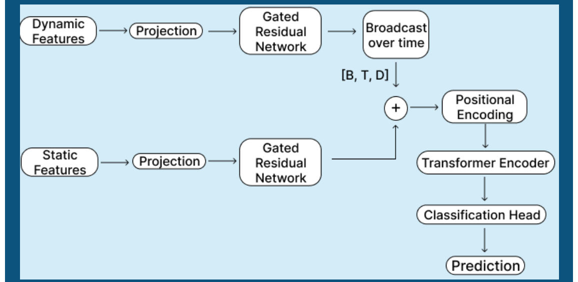
Transformer



Gated Transformer Network



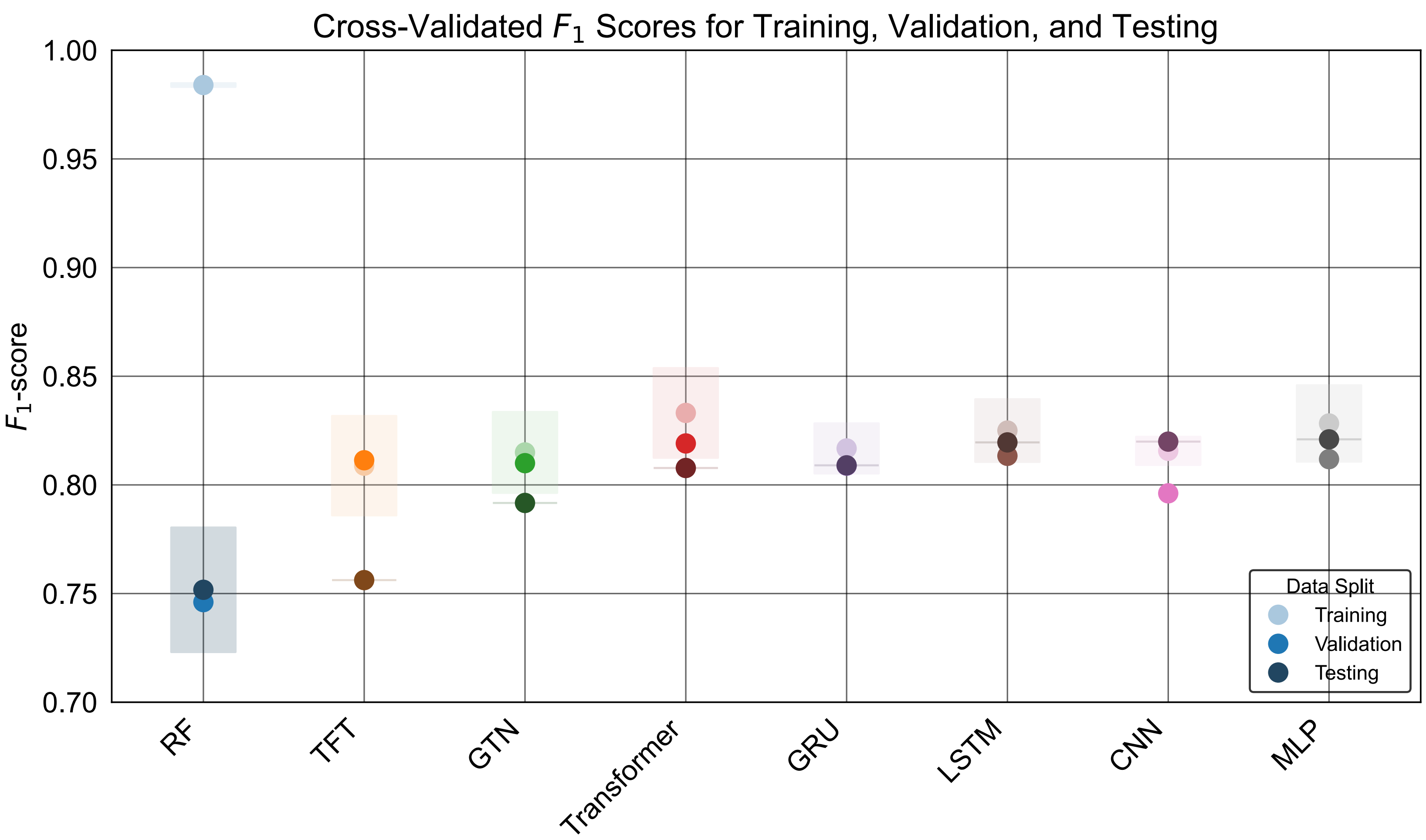
Temporal Fusion Transformer (TFT)



Performance on Wildfire Danger Prediction

All models achieved strong and robust performance with F_1-Scores > 0.75 on the test set

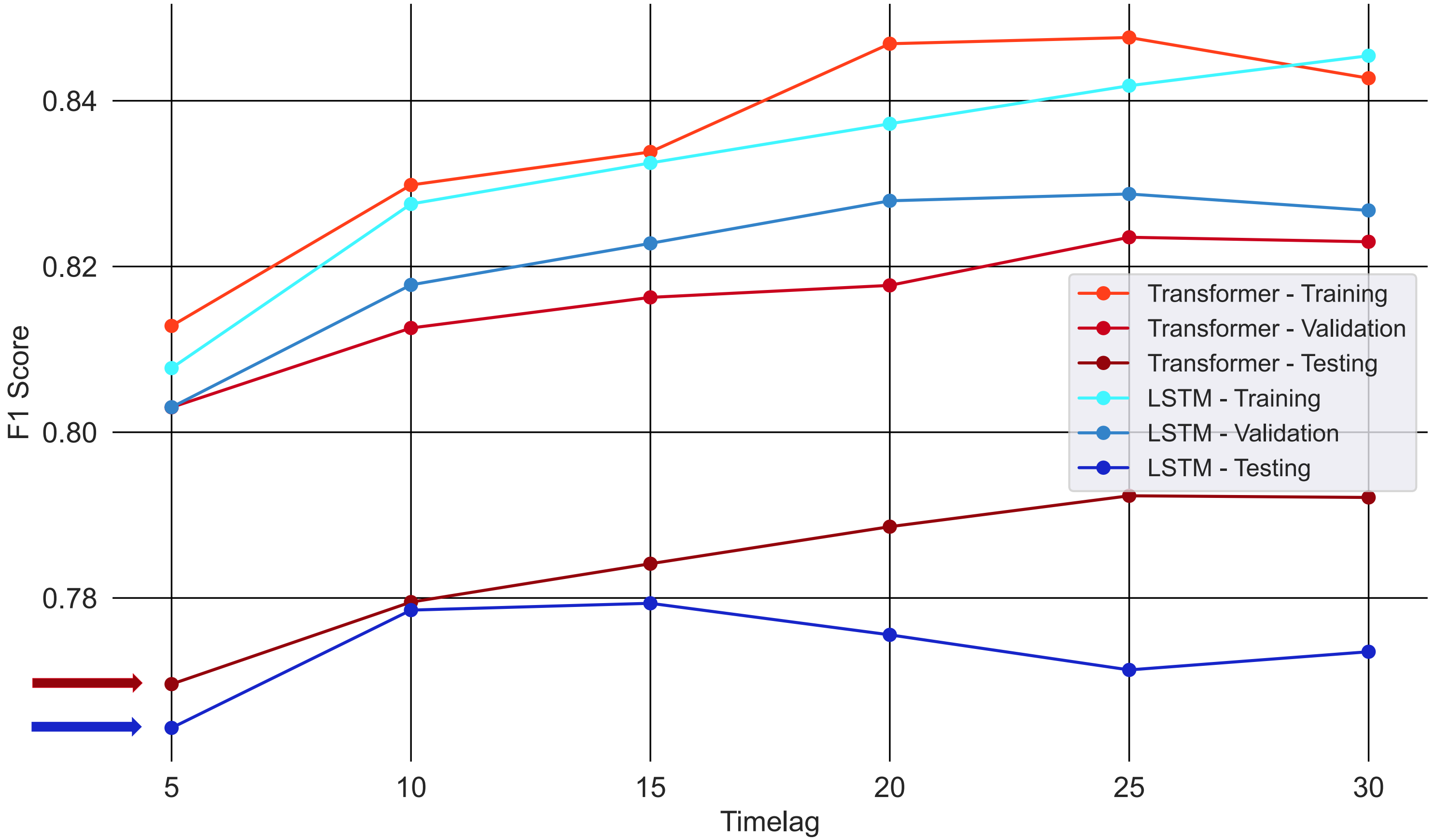
- Deep learning models consistently outperformed the RF baseline
- RF showed notably more overfitting than all other DL models
- More complex Transformers (TFT, GTN) showed no significant improvement over Transformer
- Statistical tests (ANOVA, Kruskal–Wallis) confirmed significant model differences,
  but only RF vs. deep learning models (Post-hoc Dunn test)



Impact of Time lag on Model Performance

Transformer and LSTM achieve high F_1-Scores across all time lags

- Transformer consistently outperforms the LSTM across all input lengths
- Attention mechanisms enable Transformers to better capture long-range dependencies
- Transformer benefits from longer temporal contexts
- LSTMs performance drops over 15 days



AI Climate

Fire Danger Softmax Probabilities

All testing samples (2021-2022), Transformer

- The dataset is overall imbalanced, with two-thirds negative samples
  and one-third positive samples.

- High fire probabilities (red) are concentrated along Mediterranean coastal regions, aligning with known fire-prone zones.

- Inland areas of Spain and North Africa predominantly display low fire probabilities (blue).

Positive Fire Samples in the Mediterranean

Geographic imbalance limits model capabilities, making inland fire scenarios harder to predict.

- Map shows all actual Fires (Positive Samples) in Testing Data (2021-2022)
- Inland areas show low fire probabilities (blue) → False Negatives
- **Geographic imbalance:** 72.5% of fire events are located near the coast

Explainable AI Methods

SHAP

Integrated Gradients

Physical Consistency Scores

Random Forest Feature Importance

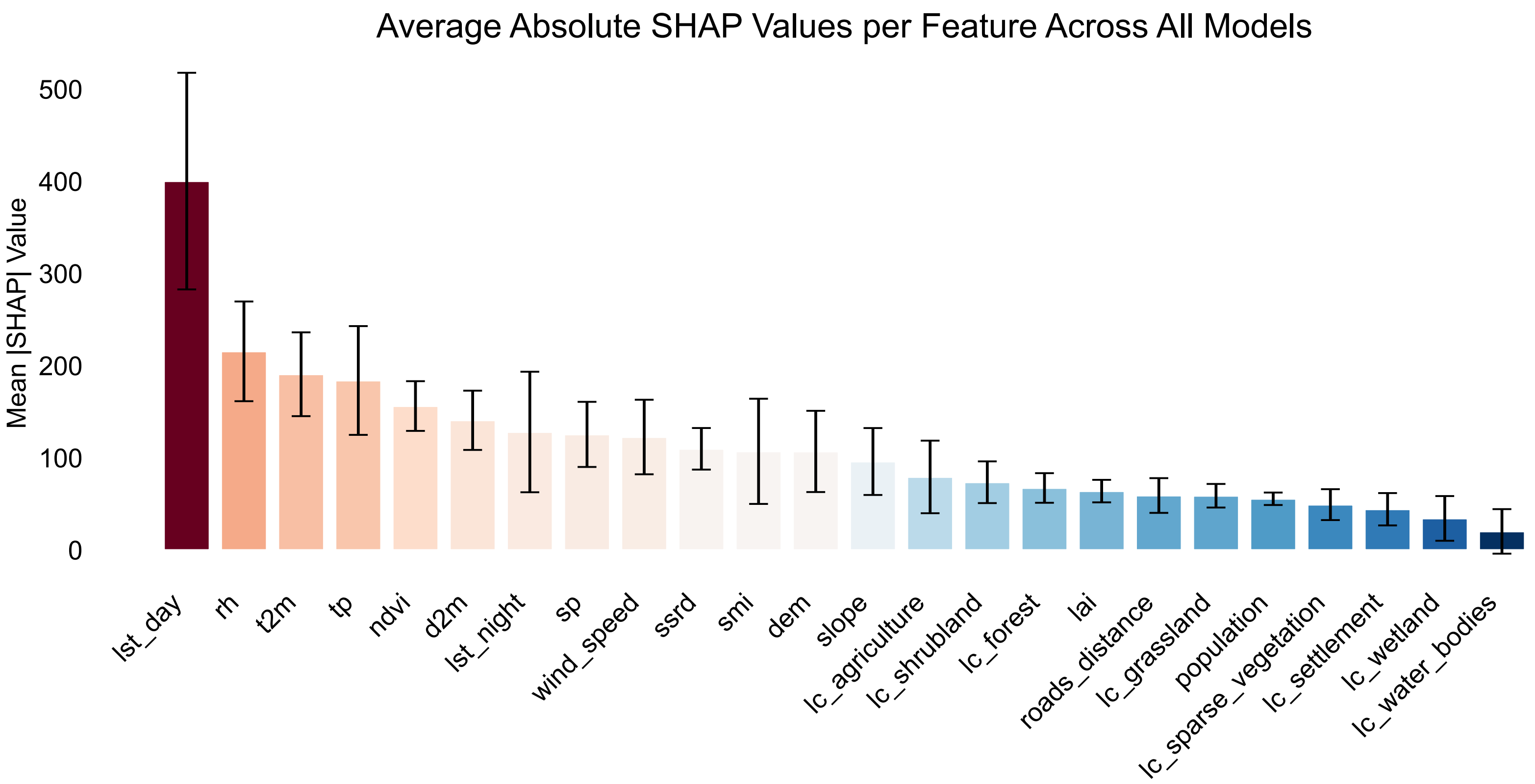
Accumulated Local Effects

Correlation Analysis

Comparison of Physical Interpretability Across Machine Learning Architectures

Average Absolute SHAP Values per Feature

- **Land day Surface Temperature:**
  Important feature across all models
- **Meteorology Variables:**
  High Absolute SHAP Values
  → Explainable Variance
- **Land cover categories (Fuel):**
  show lower SHAP values
- **Topographical Factors** and **Human Factors**:
  show medium to low SHAP values



Comparison of Physical Interpretability Across Machine Learning Architectures

Binary Physical Consistency Scores Matrix

- Compares SHAP value direction against expected physical relationships
- A feature is marked "physically consistent" if its SHAP value aligns with the expected sign
- "+" means a variable increases fire likelihood
- "–" means it decreases it

	d2m	lc_agriculture	lc_forest	lc_grassland	lc_settlement	lc_shrubland	lc_sparse_vegetation	lc_water_bodies	lc_wetland	lst_day	lst_night	rh	roads_distance	slope	smi	ssrd	t2m	tp	wind_speed
MLP	N	N	N	Y	Y	N	N	N	N	Y	Y	Y	N	Y	Y	N	Y	Y	N
LSTM	Y	N	N	Y	N	N	N	N	N	Y	Y	N	N	Y	Y	Y	Y	Y	N
GRU	Y	N	N	Y	N	N	N	Y	N	Y	Y	N	N	Y	Y	Y	Y	Y	N
RF	N	N	Y	Y	Y	N	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
Transformer	Y	N	Y	Y	N	N	N	N	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
GTN	N	Y	Y	Y	N	N	N	N	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
TFT	N	N	Y	Y	N	N	N	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
CNN	N	Y	Y	Y	N	N	N	N	N	Y	N	Y	N	Y	Y	Y	Y	Y	N
Physical-knowledge	-	+	+	+	-	+	+	-	-	+	+	-	+	+	-	+	+	-	+

Continuous Physical Consistency Scores Matrix

- Score = proportion of samples where SHAP aligns with expected sign
- Temperature variables, slope, smi and participation aligned strongly with fire-promoting effects
- Land cover features showed low consistency

	d2m (-)	lc_agriculture (+)	lc_forest (+)	lc_grassland (+)	lc_settlement (-)	lc_shrubland (+)	lc_sparse_vegetation (+)	lc_water_bodies (-)	lc_wetland (-)	lst_day (+)	lst_night (+)	rh (-)	roads_distance (+)	slope (+)	smi (-)	ssrd (+)	t2m (+)	tp (-)	wind_speed (+)
cnn	0.38	0.65	0.69	0.61	0.48	0.44	0.45	0.47	0.00	0.91	0.47	0.65	0.50	0.63	0.81	0.64	0.76	0.73	0.44
mlp	0.48	0.27	0.44	0.60	0.50	0.38	0.43	0.01	0.44	0.93	0.64	0.51	0.46	0.73	0.71	0.49	0.65	0.80	0.42
gru	0.63	0.25	0.46	0.58	0.49	0.33	0.44	0.51	0.00	0.93	0.68	0.45	0.40	0.69	0.68	0.61	0.78	0.78	0.43
lstm	0.70	0.19	0.34	0.58	0.48	0.27	0.41	0.01	0.00	0.93	0.71	0.44	0.38	0.64	0.71	0.53	0.81	0.78	0.45
transformer	0.59	0.47	0.58	0.59	0.49	0.35	0.50	0.01	0.43	0.93	0.73	0.60	0.37	0.66	0.72	0.64	0.66	0.75	0.34
gtn	0.43	0.52	0.76	0.62	0.46	0.49	0.48	0.47	0.46	0.93	0.69	0.63	0.43	0.66	0.68	0.59	0.64	0.77	0.46
tft	0.43	0.42	0.59	0.61	0.49	0.30	0.48	0.55	0.50	0.94	0.77	0.63	0.39	0.83	0.79	0.61	0.74	0.80	0.46

Consistency Score: 0.0 – 1.0

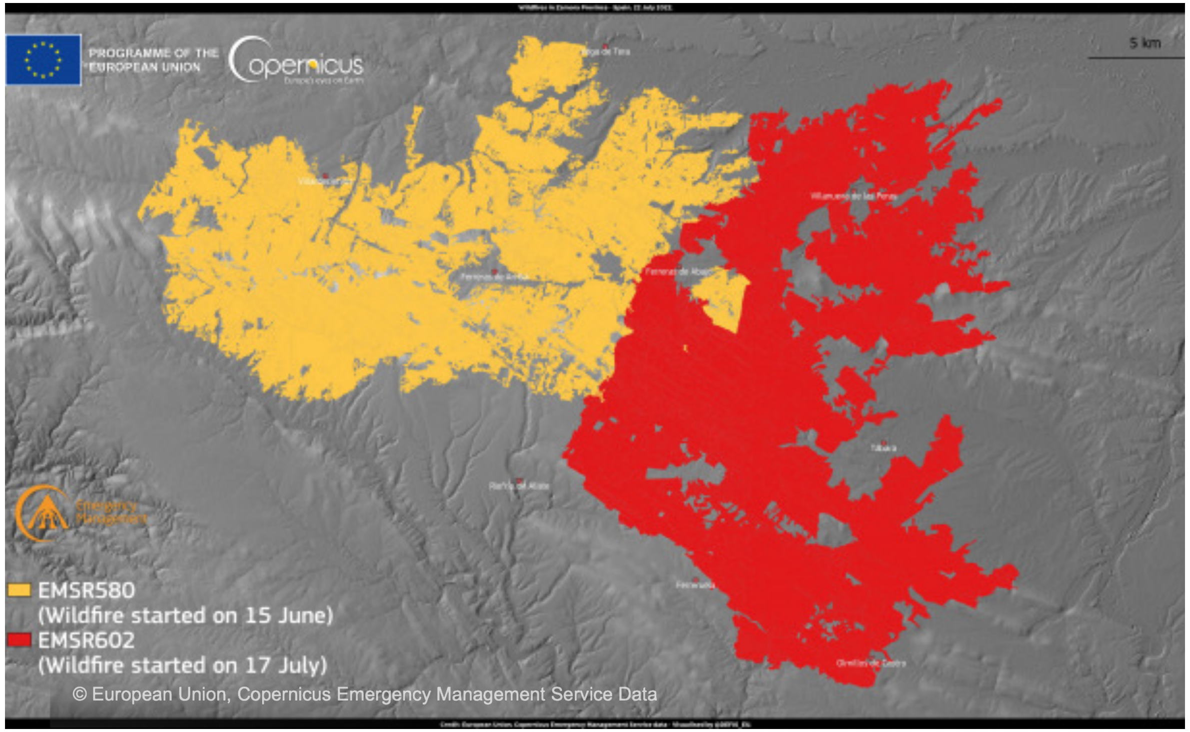
Comparison of Physical Interpretability Across Machine Learning Architectures

Binary Physical Consistency Scores Matrix

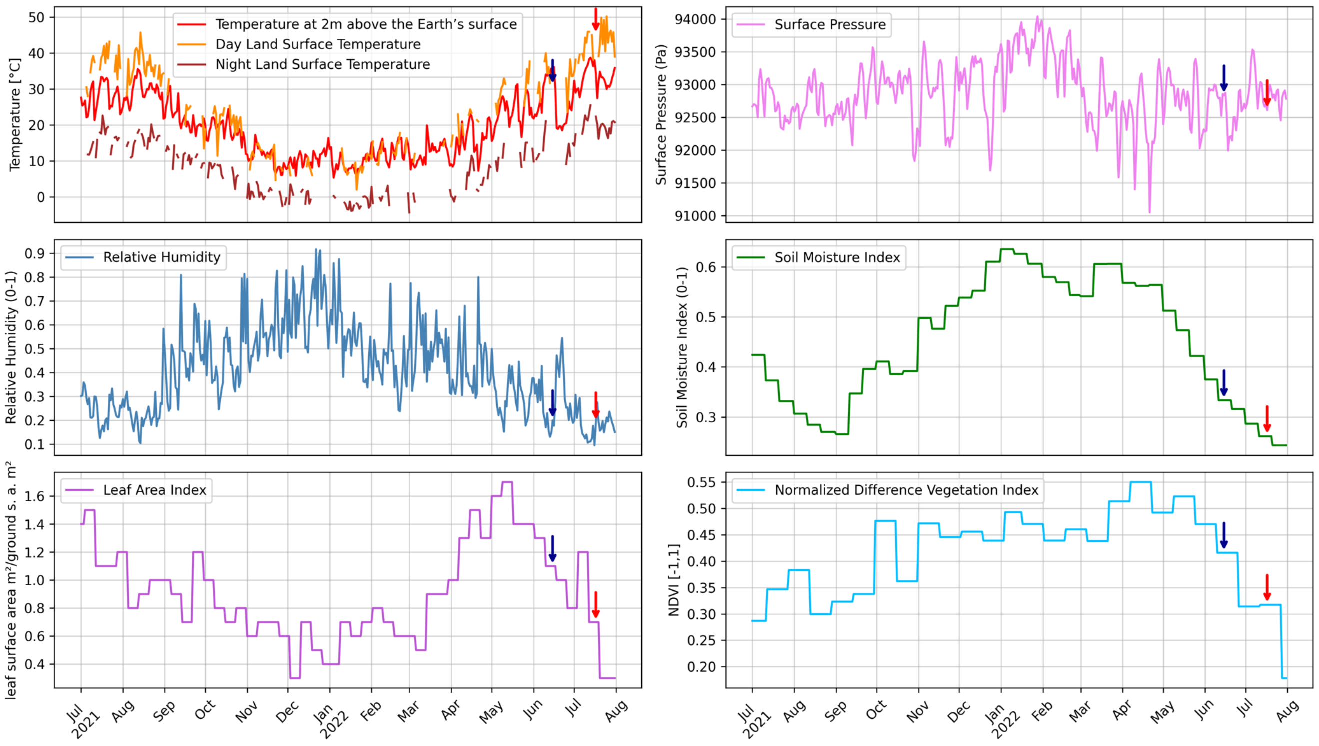
- **Transformer & GTN:** Highest F_1-Score, but only captured 11/19 expected relationships
- **Random Forest:** Highest physical consistency with 13/19 correctly predicted relationships
- **Trade-Off:** Accuracy vs. Interpretability

	d2m	lc_agriculture	lc_forest	lc_grassland	lc_settlement	lc_shrubland	lc_sparse_vegetation	lc_water_bodies	lc_wetland	lst_day	lst_night	rh	roads_distance	slope	smi	ssrd	t2m	tp	wind_speed
MLP	N	N	N	Y	Y	N	N	N	N	Y	Y	Y	N	Y	Y	N	Y	Y	N
LSTM	Y	N	N	Y	N	N	N	N	N	Y	Y	N	N	Y	Y	Y	Y	Y	N
GRU	Y	N	N	Y	N	N	N	Y	N	Y	Y	N	N	Y	Y	Y	Y	Y	N
RF	N	N	Y	Y	Y	N	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
Transformer	Y	N	Y	Y	N	N	N	N	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
GTN	N	Y	Y	Y	N	N	N	N	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
TFT	N	N	Y	Y	N	N	N	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	N
CNN	N	Y	Y	Y	N	N	N	N	N	Y	N	Y	N	Y	Y	Y	Y	Y	N
Physical-knowledge	-	+	+	+	-	+	+	-	-	+	+	-	+	+	-	+	+	-	+

Case Studies: Two Wildfires in Spain 2022



Burned Areas of Two Big Wildfires in Zamora Province, Spain in June and July 2022

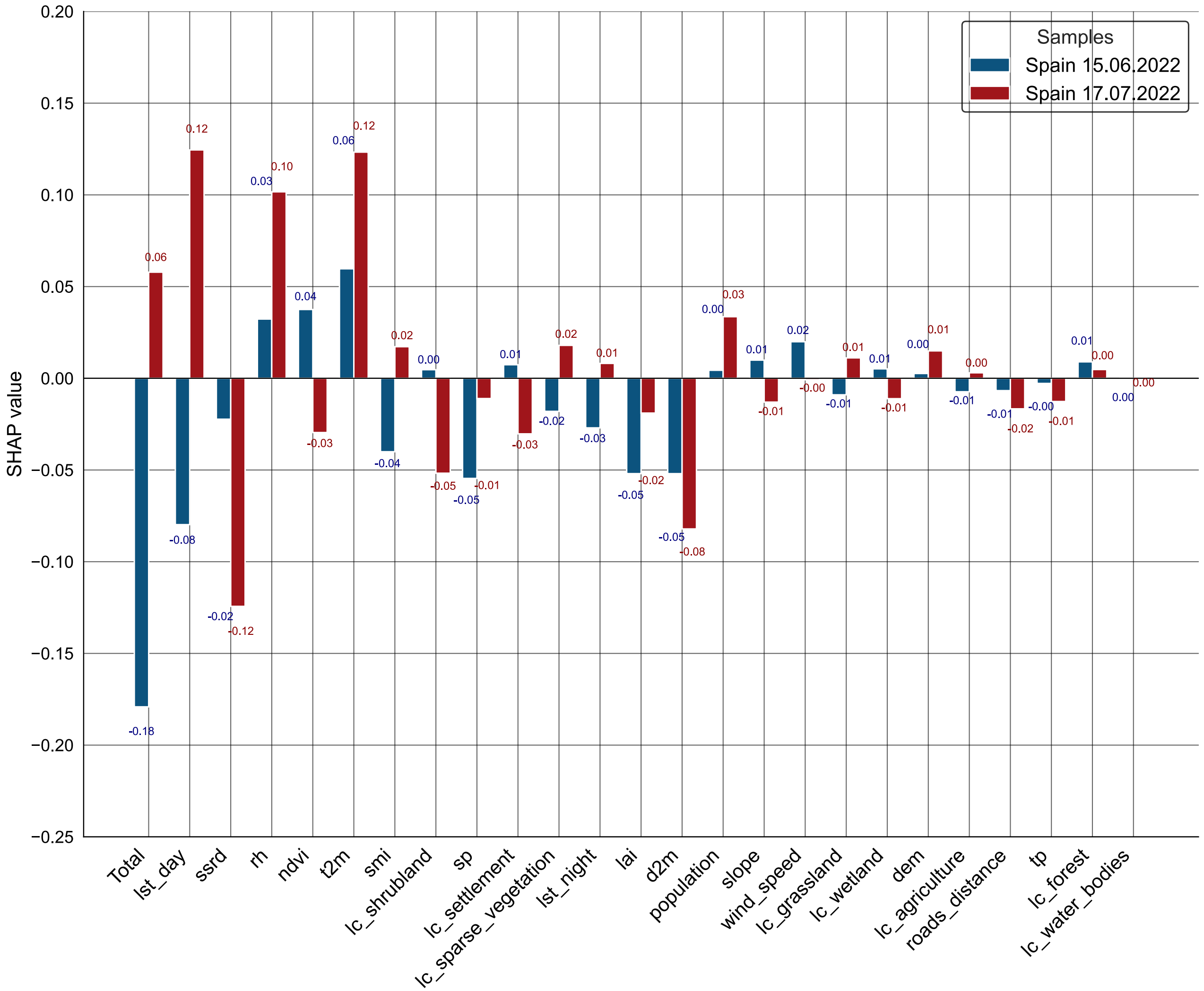


Temporal Evolution of Environmental Variables before the two Spain Fires in June and July 2022

Case Studies: Two Wildfires in Spain 2022

False Negative June vs. True Positive July

- June fire (False Negative):
  - Suppressed fire probability due to moderate temperatures and higher humidity
  - despite actual ignition by lightning
- July fire (True Positive):
  - Strong positive SHAP contributions
  - linked to extreme heatwave and drought conditions
- Key differentiators:
  - Surface and air temperatures (lst_day, t2m)
  - humidity (rh), ssrd
- Missing ignition features: Lack of lightning and ignition-related data limits model detection of lightning-induced fires.



Case Studies: Training without *day land temperature*

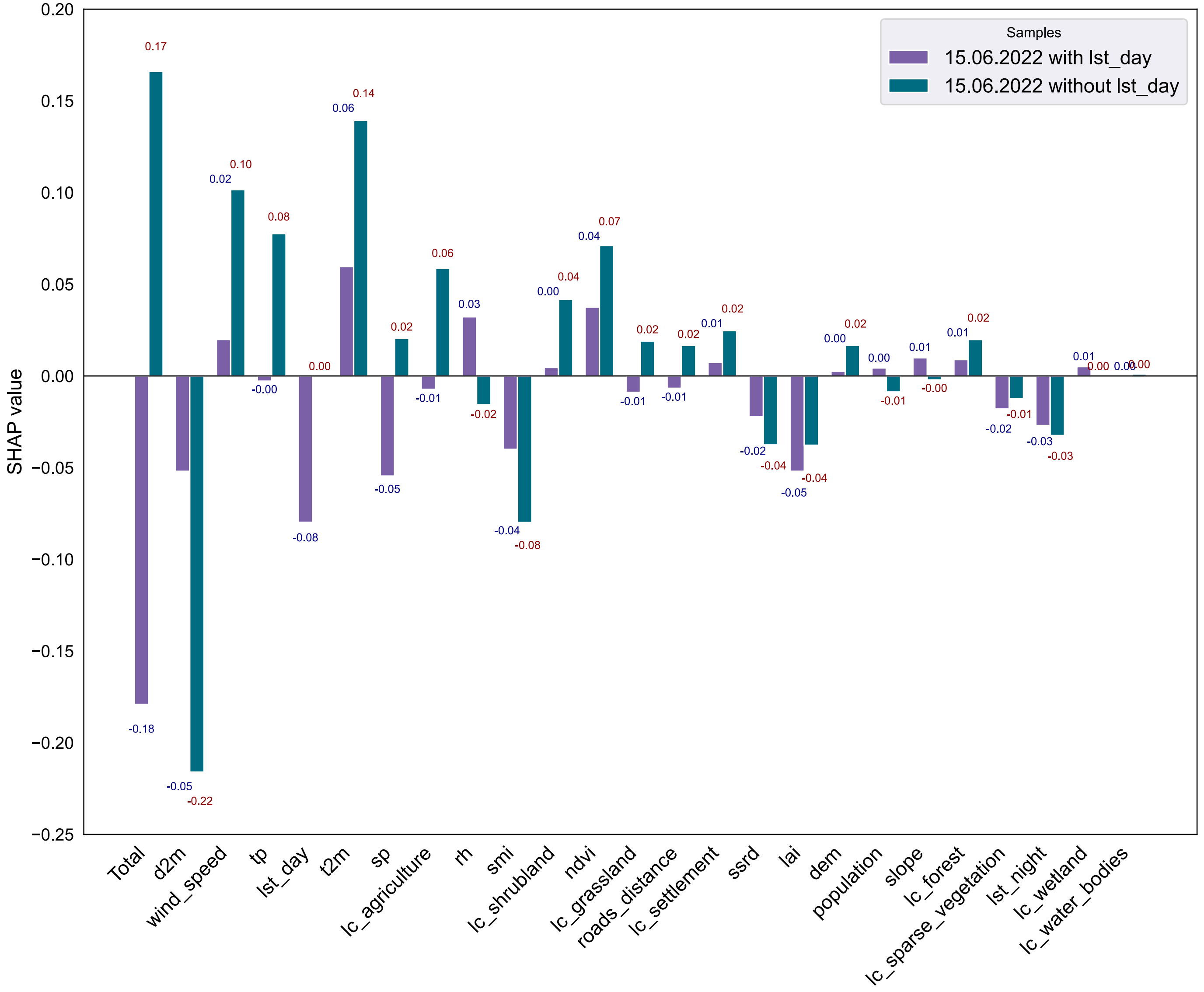
June Fire with lst_day vs. June Fire without lst_day

Effect of excluding lst_day:

- Increases fire probability from 0.19 to 0.52
- → Correct classification of June fire

SHAP redistributions:

- Stronger positive contributions from t2m, tp, ndvi, and wind speed
- lai and agriculture shift from strong negative to weaker or slightly positive attribution



AI Climate

Conclusion & Outlook

Conclusion

- All Deep Learning Models **high predictive performance**
- **No** major **accuracy gain** with increasing **model complexity**
- **Transformer** capturing best **Long-Term Dependencies**

- Model **bias** towards **costal regions**

- **Trade-Off** between **forecasting accuracy** and **scientific interpretability**

- **Multicollinearity** among temperature-related variables did not degrade model performance, but introduced **interpretability challenges** → Case Studies

Outlook

- Further Analysis of False Negatives: More **generalized analysis** of **misclassified samples** e.g. Clustering Analysis
- **Incorporating additional wildfire drivers** such as lightning strikes and human activity

- **Coastal** vs. **Inland Fire**

- Bridging Accuracy & Interpretability: **Hybrid models** (=Physical Models + DL)

- Need for multi-method explainability or **specific XAI** methods to ensure consistent feature importance even when input variables are **correlated**

Contact

Pauline Becker

Bachelor student

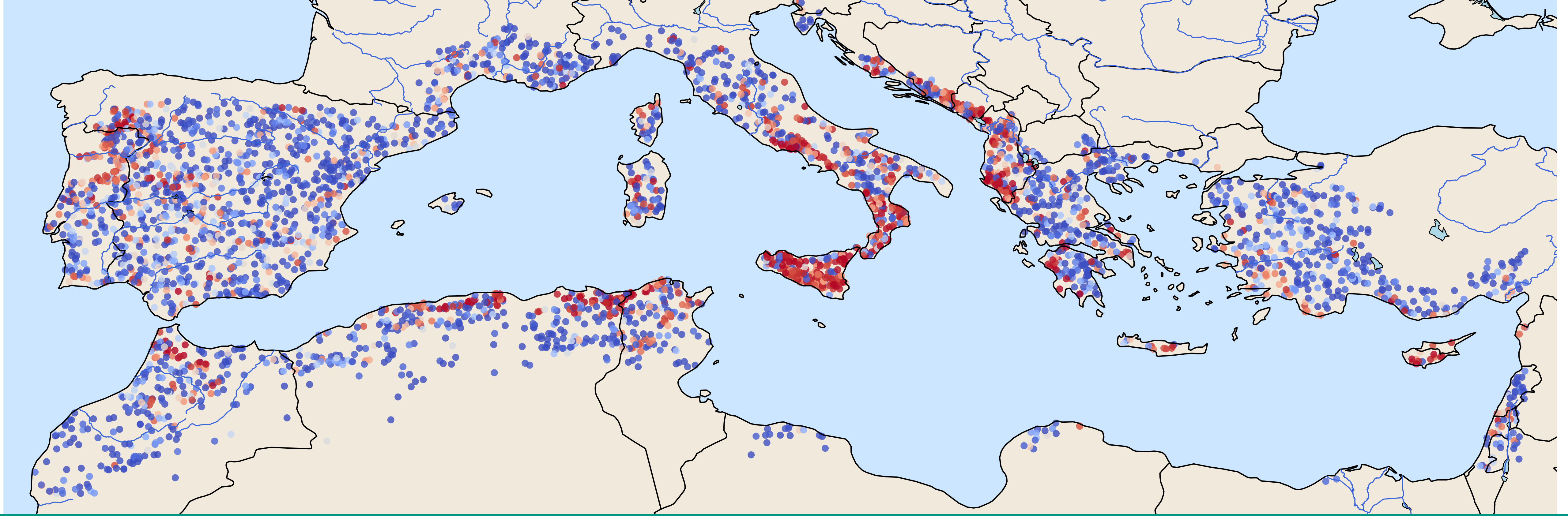
+49 152 07872365

pauline.becker@student.kit.edu

Karlsruher Institut für Technologie

Adenauerring 2
76131 Karlsruhe
Germany

KIT



Appendix

Fire Danger Softmax Probabilities

Positive Samples only

- Map shows all actual Fires in Testing Data (2021-2022)
- 72.5% of positive fire events are located near the coast
- Inland areas show low fire probabilities (blue) → False Negatives

Negative Samples only

- The dataset is overall imbalanced, with roughly two-thirds negative samples and one-third positive samples.
- Geographic imbalance limits model generalization, making inland fire scenarios harder to predict.