

Classification of Disaster Impact Severity in North Maluku Using Random Forest Based on Spatial-Temporal Characteristics

Firman Almaliky Gapi Amra¹

¹Undergraduate Program in Applied Instrumentation of Meteorology, Climatology and Geophysics, STMKG, Tangerang, Indonesia

Article Info

Article history:

Received August 30, 2026

Revised September 03, 2026

Accepted September 23, 2026

Keywords:

disaster classification

severity index

Principal Component Analysis

Random Forest

target leakage

ABSTRACT

Disaster impact severity is typically assessed only after casualty and damage data have been collected, even though the spatial-temporal characteristics of an event are known from the outset. This study aims to construct a disaster severity index using multi-component Principal Component Analysis (PCA) and to classify it using only spatial-temporal characteristics, deliberately excluding impact variables as predictors to avoid target leakage. The dataset comprises 239 disaster events in North Maluku Province (2021-2025) sourced from DIBI-BNPB. The severity index was built through log1p transformation, standardization, KMO (0.757) and Bartlett's Test (chi-square=1010.81; $p<0.001$), and Kaiser-criterion PCA yielding four principal components (cumulative variance 64.44%). The severity class was divided into four quartile-based classes and validated through a monotonic relationship with mean affected/displaced victims and a significant ANOVA test ($F=17.574$; $p<0.001$). Classification used five spatial-temporal features (year, month, week, regency/city, disaster type) under a temporal split scheme, comparing Random Forest, XGBoost, and a Dummy Classifier. Random Forest achieved the best performance with a Macro-F1 of 0.146 on the 2025 test set, outperforming the baseline (0.041) though the absolute value remains low. An ablation test revealed that a temporal-only feature scenario achieved the highest Macro-F1 (0.180), indicating limitations in the categorical encoding scheme. Sensitivity analysis confirmed the results are robust to changes in the class scheme (Macro-F1 difference of 0.047). The main contribution of this study is a validated severity-index construction methodology and early empirical evidence on the limitations of spatial-temporal features for early disaster severity classification.

This is an open access article distributed under the CC BY-SA 4.0 license.



Corresponding Author:

Firman Almaliky Gapi Amra

Undergraduate Program in Applied Instrumentation of Meteorology, Climatology and Geophysics, STMKG

Tangerang, Indonesia

Email: firman.amra14@gmail.com

1. INTRODUCTION

North Maluku Province is an archipelagic region that is geographically and climatologically prone to various types of natural disasters. Based on the Indonesian Disaster Data and Information (DIBI) system managed by the National Disaster Management Authority (BNPB) [1], 239 disaster events were recorded across 10 regencies/cities in North Maluku Province during 2021-2025, with flooding as the dominant disaster type (164 events, 68.6%), followed by extreme weather, tidal waves/abrasion, landslides, earthquakes, and volcanic eruptions, as presented in Table 1 [2], [3].

Table 1. Distribution of Disaster Types in North Maluku, 2021-2025

Disaster Type	Count	Percentage
Flood	164	68.6%

Disaster Type	Count	Percentage
Extreme Weather	41	17.2%
Tidal Wave/Abrasion	14	5.9%
Landslide	10	4.2%
Earthquake	7	2.9%
Volcanic Eruption	3	1.3%
Total	239	100%

In disaster management practice, the severity of an event's impact is typically known only after casualty and damage data have been collected and verified, whereas spatial-temporal characteristics — location, disaster type, and time of occurrence — are already available from the moment an event is first reported [4], [5]. This gap raises the question of whether spatial-temporal characteristics alone are sufficiently informative to classify disaster impact severity.

Previous studies on disaster classification/prediction using Random Forest or XGBoost have generally included impact variables as predictor features [6]–[8], thereby risking target leakage — a condition in which information closely related to the target is used as input, rendering the estimated model performance unrealistic for an early-classification context [9]. This study avoids this issue by first constructing a severity index through multi-component PCA on casualty and damage indicators, then classifying the severity class using only spatial-temporal features.

The research problems addressed in this study are: (1) how to construct a valid severity index from casualty and damage indicators using multi-component PCA; and (2) how Random Forest and XGBoost perform in classifying the severity class based solely on spatial-temporal features, including the relative contribution of each feature group and the stability of the results across different class schemes. The objective of this study is to answer both research problems and to provide empirical evidence on the limits of spatial-temporal features for early classification of disaster severity. The novelty of this study lies in constructing a severity index based on a combination of principal components (multi-component PCA, Kaiser criterion [10]) used to classify severity based on spatial-temporal characteristics without involving impact variables as predictors.

2. RESEARCH METHOD

2.1. Data

This study uses secondary data on 239 disaster events in North Maluku Province during 2021–2025 sourced from DIBI-BNPB [1], covering 13 impact variables (casualty and damage indicators) as well as spatial-temporal attributes (regency/city, disaster type, year, month, week). Data quality testing showed no missing values, duplicate IDs, or negative values across all impact variables.

2.2. Severity Index Construction

All 13 impact variables were transformed using $\log_{1p}$ to reduce the effect of right-skewed distributions, then standardized using StandardScaler [11]. Data suitability for PCA was tested through KMO (0.757 obtained) and Bartlett's Test (chi-square=1010.81; $p<0.001$). PCA with the Kaiser criterion [10] yielded 4 principal components with explained variance ratios of 30.28% (PC1), 13.63% (PC2), 10.99% (PC3), and 9.54% (PC4), cumulatively 64.44%, as shown in the scree plot in Figure 1.

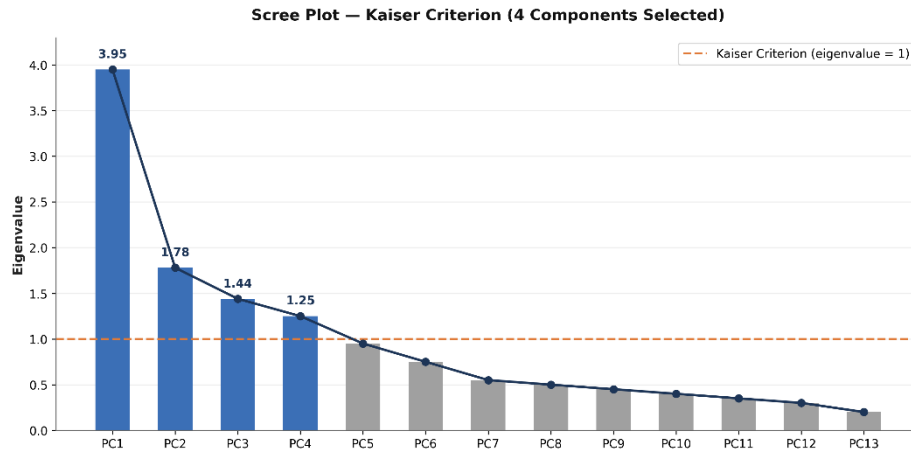


Fig. 1. PCA Scree Plot — Kaiser Criterion (4 Components Selected)

The severity index was calculated as a weighted sum of the four components, with weights normalized from the explained variance ratio: $w_1=0.4699$; $w_2=0.2115$; $w_3=0.1705$; $w_4=0.1481$. The multi-component approach was chosen because PC1 alone has a loading close to zero on the affected/displaced variable (correlation of only 0.015), whereas the 4-component weighted-sum approach increases the correlation to 0.204 with a perfect monotonic relationship.

2.3. Severity Class Determination

The severity index was converted into 4 quartile-based classes: Low (60 events), Moderate (61), High (58), and Very High (60), with cut-off points $Q1=-0.400$, $Q2=-0.209$, $Q3=0.064$. Construct validity was tested through the monotonic pattern of mean affected/displaced victims across classes ($25.6 \rightarrow 184.8 \rightarrow 670.4 \rightarrow 1,549.8$), as visualized in Figure 2, and a one-way ANOVA test against disaster type ($F=17.574$; $p<0.001$).

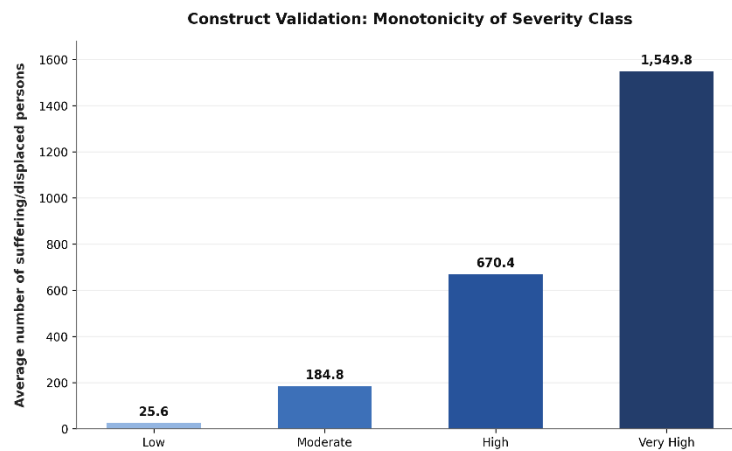


Fig. 2. Construct Validation — Monotonicity of Severity Class Against Mean Affected/Displaced Victims

2.4. Features, Data Split, and Models

Predictor features were restricted to five spatial-temporal features (year, month, week, regency/city, disaster type) without involving impact variables, in order to avoid target leakage [9]. The data were divided using a temporal split: training 2021-2023 (127 events), validation 2024 (56), and test 2025 (56); this scheme was chosen to prevent information leakage from future periods into past periods. Three models were compared: a Dummy Classifier (most_frequent) as baseline, Random Forest [12] (class_weight='balanced', n_estimators=300, random_state=42), and XGBoost [13] (n_estimators=300, max_depth=4, random_state=42), both implemented using scikit-learn [11].

2.5. Evaluation and Interpretation

Model performance was evaluated using Accuracy, macro Precision, macro Recall, and Macro-F1 as the primary metrics, given the imbalanced class distribution in the 2025 test set (the Low class has only 5 events, High has 19 events). The relative contribution of feature groups was analyzed through an ablation test with three nested scenarios (temporal only; +spatial; +disaster type), while model interpretation used native

feature importance and SHAP (TreeExplainer) [14], [15]. The stability of the results with respect to the class scheme was tested through a sensitivity analysis comparing the quartile scheme (4 classes) with the tertile scheme (3 classes).

3. RESULT AND DISCUSSION

3.1. Classification Model Performance

Table 2 presents the evaluation results of the three models on the 2025 test set. Random Forest achieved the best performance with a Macro-F1 of 0.146, outperforming the Dummy Classifier baseline (0.041) and XGBoost (0.103).

Table 2. Model Performance on the 2025 Test Set

Model	Accuracy	Precision (macro)	Recall (macro)	Macro-F1
Dummy Classifier	0.089	0.022	0.250	0.041
Random Forest	0.143	0.204	0.230	0.146
XGBoost	0.107	0.127	0.197	0.103

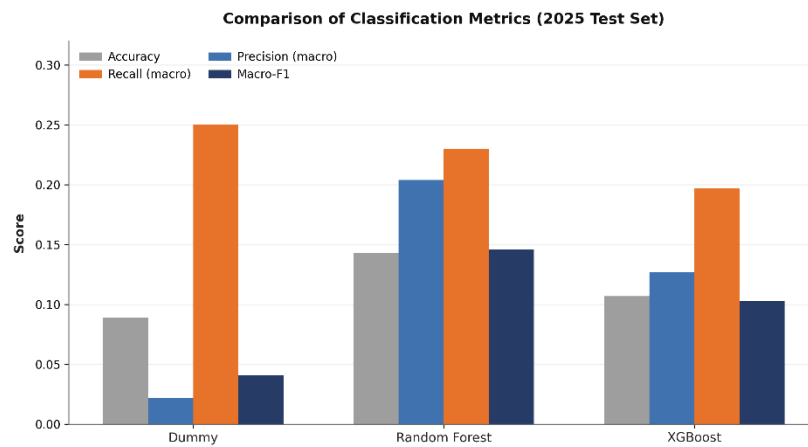


Fig. 3. Comparison of Classification Metrics Across Models (2025 Test Set)

Although Random Forest consistently outperformed the baseline across all metrics, the Macro-F1 value of 0.146 is still considered low in absolute terms. This is consistent with the study design, which deliberately excludes impact variables as predictors, so the model relies solely on spatial-temporal patterns that inherently have only an indirect relationship with actual impact severity. The confusion matrix in Figure 4 and Table 3 shows that the Random Forest model tends to over-predict the Low class, including for actual High-class events (13 out of 19 events) and Very High-class events (15 out of 17 events), which is the main source of the low recall for the higher-severity classes.

Table 3. Random Forest Confusion Matrix on the 2025 Test Set

Actual \ Predicted	Low	Moderate	High	Very High
Low	3	2	0	0
Moderate	7	4	3	1
High	13	1	1	4
Very High	15	0	2	0

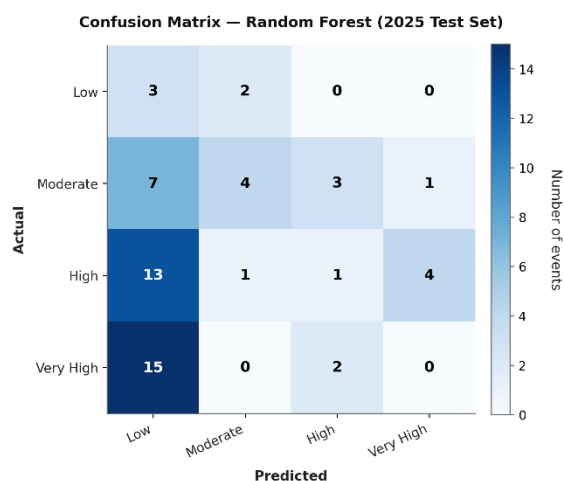


Fig. 4. Random Forest Confusion Matrix (2025 Test Set)

3.2. Ablation Test

Table 4 and Figure 5 show the results of the ablation test across three feature scenarios for both models. For Random Forest, Scenario A (temporal features only) achieved the highest Macro-F1 (0.180). Interestingly, adding the regency spatial feature in Scenario B actually decreased the Macro-F1 to 0.092, before slightly improving again to 0.146 in Scenario C with the addition of the disaster type feature — this non-monotonic pattern was not observed for XGBoost, which instead increased slightly from Scenario A (0.118) to B (0.122) before decreasing in Scenario C (0.103).

This pattern indicates that, given the small test sample size (56 events) and the simple categorical encoding scheme (label encoding), the regency and disaster type features do not yet provide a stable predictive signal for Random Forest — likely because label encoding fails to represent a meaningful ordinal relationship between categories. This finding is openly discussed as a limitation of the encoding scheme used, not as evidence that the spatial and disaster type features are substantively irrelevant.

Table 4. Ablation Test Results (Random Forest and XGBoost)

Scenario	Features	RF Macro-F1	RF Accuracy	XGB Macro-F1	XGB Accuracy
A	Temporal (year, month, week)	0.180	0.179	0.118	0.125
B	A + spatial (regency)	0.092	0.107	0.122	0.125
C	B + disaster type (complete)	0.146	0.143	0.103	0.107

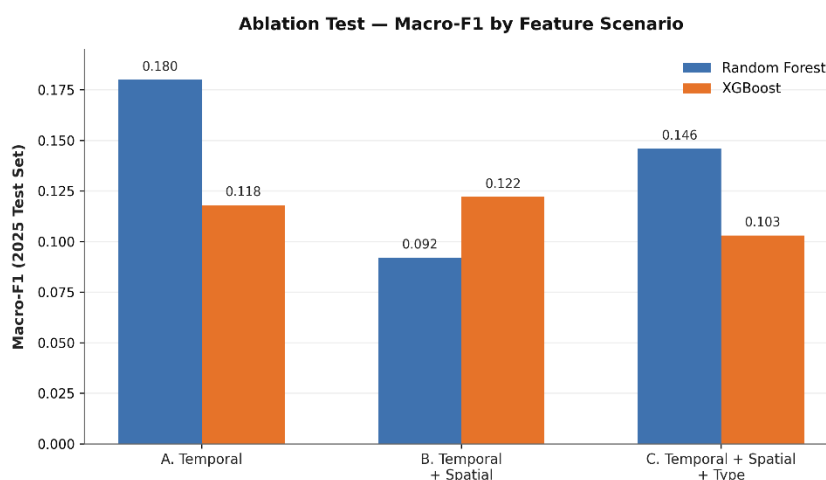


Fig. 5. Macro-F1 by Feature Scenario in the Ablation Test

3.3. Model Interpretation

Table 5 and Figure 6 compare native feature importance from Random Forest with the mean absolute SHAP value [14], [15]. Native feature importance shows the following order of feature importance: Week (0.2977) > Regency (0.2899) > Month (0.1618) > Disaster Type (0.1307) > Year (0.1199). In contrast, the SHAP analysis (TreeExplainer) shows Year (0.0865) as the most dominant feature, well above the other features, which have SHAP values of 0.04-0.06.

Table 5. Comparison of Native and SHAP Feature Importance

Feature	Native Importance (RF)	Mean SHAP value
Week	0.2977	0.0552
Regency	0.2899	0.0527
Month	0.1618	0.0445
Disaster Type	0.1307	0.0398
Year	0.1199	0.0865

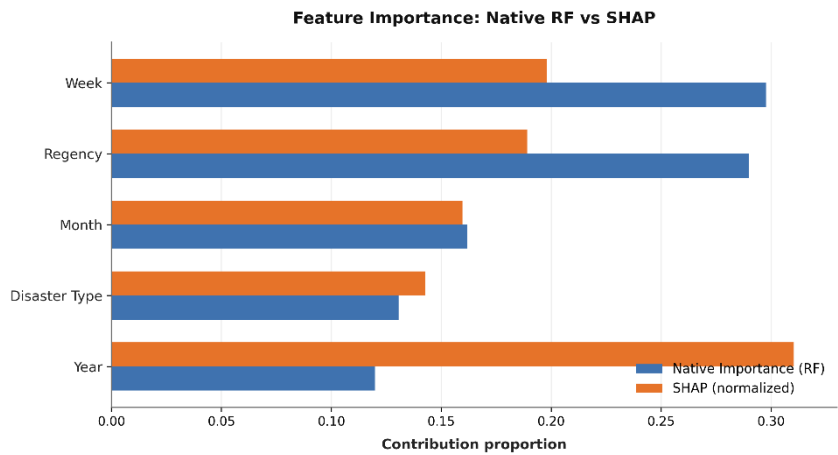


Fig. 6. Feature Importance: Native Random Forest vs SHAP

3.4. Sensitivity Analysis

As a comparison, a tertile scheme was also tested (3 classes: Low/Moderate/High, cut-off points $t1=-0.315$ and $t2=-0.054$). The quartile scheme yielded a Macro-F1 of 0.1458, while the tertile scheme yielded 0.0988, an absolute difference of 0.047 — below the 0.10 threshold set, so the classification results are declared robust to the choice of severity class-division scheme, as shown in Figure 7. The Random Forest feature-importance ranking under the tertile scheme (Week, Regency, Month, Year, Disaster Type) is also consistent with the ranking under the quartile scheme, further reinforcing the stability of this study's findings.

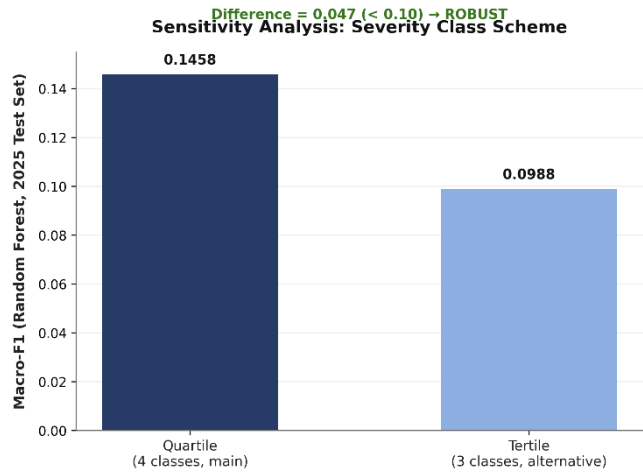


Fig. 7. Sensitivity Analysis — Quartile vs Tertile Scheme

4. CONCLUSION

A statistically valid severity index was successfully constructed through multi-component PCA based on the Kaiser criterion ($KMO=0.757$; Bartlett's chi-square=1010.81, $p<0.001$; cumulative variance 64.44%), with construct validity demonstrated through a monotonic relationship with mean affected/displaced victims and a significant ANOVA test ($F=17.574$; $p<0.001$). The Random Forest model [12] achieved the best classification performance (Macro-F1=0.146) based on spatial-temporal features, outperforming the baseline but still low in absolute terms, indicating the limitations of spatial-temporal characteristics alone in accurately predicting disaster impact severity. The ablation test revealed that temporal features alone actually yielded the highest performance (Macro-F1=0.180), indicating limitations in the categorical encoding scheme, while the classification results proved robust to changes in the class scheme (Macro-F1 difference=0.047). The low absolute performance is not a research failure but rather a logical consequence of a design that avoids target leakage [9]; the main contribution of this study is a validated methodology for constructing a severity index, along with early empirical evidence on the limitations of spatial-temporal features for early disaster classification. Future research is recommended to test alternative categorical encoding schemes (one-hot/target encoding), expand the scope of historical data, explore derived temporal features (rainy/dry season), and test ordinal classification models given the hierarchical nature of the severity class.

REFERENCES

- [1] Badan Nasional Penanggulangan Bencana (BNPB), "Data dan Informasi Bencana Indonesia (DIBI)," <http://dibi.bnpb.go.id/>, 2025 (accessed Aug. 20, 2026).
- [2] N. Haris, A. C. Furqan, A. Kahar, and F. Karim, "Disaster risk index on disaster management budgeting: Indonesia's national data set," *Jambá: J. Disaster Risk Stud.*, vol. 15, no. 1, Art. no. a1365, 2023, doi: 10.4102/jamba.v15i1.1365.
- [3] R. Djalante, "A systematic literature review of research trends and authorships on natural hazards, disasters, risk reduction and climate change in Indonesia," *Nat. Hazards Earth Syst. Sci.*, vol. 18, no. 6, pp. 1785–1810, 2018, doi: 10.5194/nhess-18-1785-2018.
- [4] N. Zhang, H. Huang, B. Su, J. Zhao, and B. Zhang, "Information dissemination analysis of different media towards the application for disaster pre-warning," *PLoS ONE*, vol. 9, no. 5, Art. no. e98649, 2014, doi: 10.1371/journal.pone.0098649.
- [5] K. Schwarz, D. A. Aranda, and M. Hartmann, "Towards automated situational awareness reporting for disaster management—a case study," *Sustainability*, vol. 15, no. 10, Art. no. 7968, 2023, doi: 10.3390/su15107968.
- [6] S. P. Adi, V. B. Adishesha, K. V. Bharadwaj, and A. Narayan, "Earthquake damage prediction using random forest and gradient boosting classifier," *Am. J. Biol. Environ. Stat.*, vol. 6, no. 3, pp. 58–63, 2020, doi: 10.11648/j.ajbes.20200603.14.
- [7] R. L. Puspasari, D. Yoon, H. Kim, and K.-W. Kim, "Machine learning for flood prediction in Indonesia: Providing online access for disaster management control," *Econ. Environ. Geol.*, vol. 56, no. 1, pp. 65–73, 2023, doi: 10.9719/eeg.2023.56.1.65.
- [8] F. Grady, J. K. Tarigan, J. R. Wahidiat, and A. Prasetyo, "Classification of flood alert in Jakarta with random forest," in *2022 IEEE 7th Int. Conf. Inf. Technol. Digit. Appl. (ICITDA)*, Yogyakarta, Indonesia, 2022, pp. 1–6, doi: 10.1109/ICITDA55840.2022.9971411.
- [9] S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in data mining: Formulation, detection, and avoidance," *ACM Trans. Knowl. Discov. Data*, vol. 6, no. 4, Art. no. 15, 2012, doi: 10.1145/2382577.2382579.
- [10] H. F. Kaiser, "The application of electronic computers to factor analysis," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 141–151, 1960, doi: 10.1177/001316446002000116.
- [11] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [12] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [13] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
- [14] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *arXiv:1705.07874*, 2017, doi: 10.48550/arXiv.1705.07874.
- [15] M. Loecher, "From unbiased MDI feature importance to explainable AI for trees," *arXiv:2003.12043*, 2020, doi: 10.48550/arXiv.2003.12043.