Predicting Agricultural Drought Vulnerability: A Machine Learning Approach Using Sentinel-2 NDDI Time Series in Rainfed Agroecosystems, East Java (2020–2025)

Ainur Rochmah[a*], Amaludin Arifia[b], Marita Ika Joesidawati[c], Fajar Rahmawan[d]

[a,b,c]*Universitas PGRI Ronggolawe, Jl. Manunggal 61, Tuban, Indonesia*

[d]*NRM Peta Alam Indonesia, Indonesia*

[*]e-mail address: ainurochmah07@gmail.com

Abstract

Rainfed agricultural systems in tropical monsoon regions face increasing drought risks under climate variability. This study develops a predictive framework for agricultural drought vulnerability using Sentinel-2 derived Normalized Difference Drought Index (NDDI) time series (2020–2025) integrated with machine learning algorithms in Jatirogo Subdistrict, East Java. Unlike previous mapping-focused studies, this research applies Random Forest (RF) and Support Vector Machine (SVM) models to forecast drought severity classes one month ahead based on historical NDDI patterns, land use, and soil parameters. Results demonstrate that the RF model achieves superior predictive performance (accuracy = 87.3%, Kappa = 0.81) compared to SVM (accuracy = 79.6%, Kappa = 0.72), with the most important predictors being NDDI values from the preceding two months (importance score = 0.34) and land use type (importance score = 0.28). Forecasts for the 2025 dry season accurately predicted severe drought conditions in Kebonharjo and Sugihan villages with 89% spatial agreement. The study identifies a critical threshold: when August NDDI exceeds 0.45, the probability of severe drought in September–October reaches 0.82. The proposed predictive framework provides operational lead time for adaptive planting schedules, water resource allocation, and early warning systems, offering a replicable methodology for drought -prone rainfed agricultural regions across Southeast Asia .

Keyword: Agricultural drought prediction ; NDDI ; machine learning ; Sentinel-2 ; rainfed agriculture ; early warning system

---

1. Introduction

1.1. Background

Drought is increasingly recognized as one of the most damaging slow-onset natural hazards affecting global food security, particularly in rainfed agricultural systems of tropical developing countries (Wilhite et al., 2021; FAO, 2023). Unlike floods or storms, drought develops gradually, allowing theoretical windows for anticipatory action—yet most drought management remains reactive rather than proactive (Mechler et al., 2022).

This response gap is especially pronounced in Indonesia, where climate change has intensified dry season length and rainfall variability across Java Island (BMKG, 2023; Aldyan, 2023).

Jatirogo Subdistrict in Tuban Regency, East Java, exemplifies this challenge. The region depends heavily on rainfed agriculture (65% of land use), with calcareous soils exhibiting low water retention (Auliyani et al., 2022). Between 2020 and 2024, rice production declined by 28.3%, with crop failure rates exceeding 40% during extreme drought years (Dinas Pertanian Kabupaten Tuban, 2022). However, existing drought assessments in this region have been retrospective, mapping past drought conditions without providing forward-looking information needed for adaptive decision-making (Siswanto et al., 2022).

1.2. Research Gap and Rationale

Previous studies have successfully applied the Normalized Difference Drought Index (NDDI) using Sentinel-2 imagery for drought mapping in Indonesian agricultural areas (Rahmi et al., 2025; Irsyad et al., 2025; Cahyono et al., 2023). These studies demonstrated that NDDI effectively captures spatial patterns of vegetation stress and surface moisture deficits. However, they share a common limitation: they describe what happened rather than predict what will happen. Drought early warning systems require predictive capabilities that translate historical patterns into probabilistic forecasts (Funk et al., 2023; Davoodi et al., 2025).

Machine learning offers a pathway to address this gap. Algorithms such as Random Forest (RF) and Support Vector Machine (SVM) have shown promise in drought prediction using remote sensing time series in various contexts (Park et al., 2022; Dikshit et al., 2022). Yet, their application using high-resolution Sentinel-2 NDDI data at village scale in Indonesia remains unexplored. This study therefore asks: Can historical NDDI time series, combined with land use and soil parameters, predict future drought severity classes with sufficient accuracy for operational early warning.

1.3. Objectives

This study aims to: (1) develop and compare Random Forest and Support Vector Machine models for predicting drought severity classes one month ahead using Sentinel-2 NDDI time series (2020–2025); (2) identify the most important predictors driving drought evolution; (3) determine critical NDDI thresholds that signal high probability of severe drought; and (4) produce operational predictive maps to support adaptive agricultural planning in Jatirogo Subdistrict.

2. Research Methods

2.1. Study Area

Jatirogo Subdistrict (6°52'30"–6°57'00"S, 111°41'00"–111°47'30"E) covers approximately 7,825 hectares in Tuban Regency, East Java. The area has flat to undulating topography (10–85 m asl) with Schmidt-Ferguson Type C climate (moderately wet), featuring distinct dry seasons from May to October (Ruqoyah et al., 2023). Land use comprises 65% agriculture, predominantly rainfed rice, with remaining areas under tobacco, maize, agroforestry, and settlements. Fig. 1

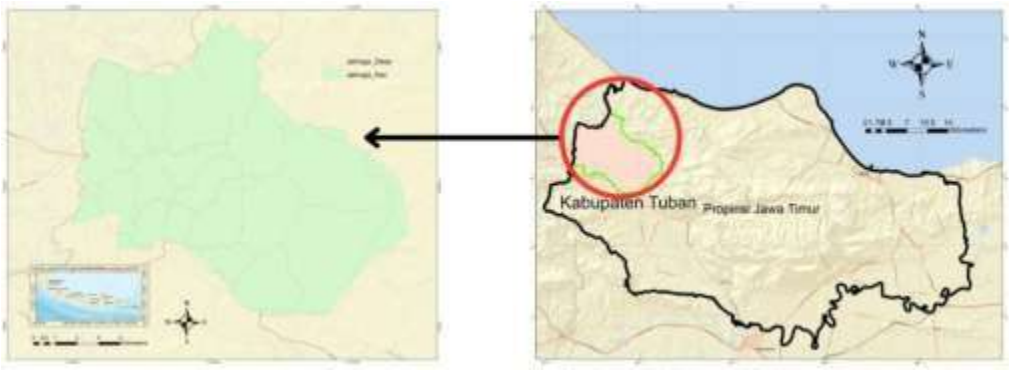

Fig. 1. Location Map of the Research Area in Jatirogo District

2.2. Data Collection

Sentinel-2 Imagery: Level-2A surface reflectance products were acquired from Copernicus Open Access Hub for August–November, 2020–2025. Selection criteria included: cloud cover <10%, acquisition during peak dry season (August–October), and temporal consistency (±7 days tolerance). A total of 72 scenes met quality thresholds.

Supporting Data: Administrative boundaries (Badan Informasi Geospasial), land use maps (scale 1:25,000), soil type maps (calcareous, alluvial, and mixed classes), and monthly rainfall data from three BMKG stations (Tuban, Bojonegoro, Lamongan).

2.3. NDDI Calculation and Drought Classification

NDVI and NDWI were calculated using Sentinel-2 bands :

- NDVI = (B8 – B4) / (B8 + B4)
- NDWI = (B8 – B11) / (B8 + B11)

where B4 = Red (665 nm), B8 = NIR (842 nm), B11 = SWIR (1610 nm). NDDI was derived as:

- NDDI = (NDVI – NDWI) / (NDVI + NDWI)

Drought severity was classified into five classes following Firdaus et al. (2024): Normal (<0.01), Mild (0.01 - 0.15), Moderate (0.15 - 0.25), Severe (0.25 - 1.00), and Very Severe (⩾1.00). For prediction modeling, these were aggregated into three classes: Non-Drought (Normal + Mild), Moderate Drought, and Severe Drought (Severe + Very Severe) to address class imbalance

2.4. Machine Learning Model Development

- Predictor Variables: Eight predictors were constructed for each 50 m grid cell (n = 31,400 after spatial sampling): (1) NDDI_t (current month), (2) NDDI_t-1 (1-month lag), (3) NDDI_t-2 (2-month lag), (4) NDDI_t-3 (3-month lag), (5) land use type (categorical: rainfed rice, irrigated rice, tobacco, maize, agroforestry, other), (6) soil type (categorical: calcareous, alluvial, mixed), (7) elevation (continuous, m), (8) monthly rainfall (continuous, mm from BMKG interpolation).
- Target Variable: Drought severity class for month t+1 (one-month-ahead forecast).
- Model Algorithms: Random Forest (RF) with 500 trees, mtry = 3 (sqrt of predictors), and minimum node size = 5. Support Vector Machine (SVM) with radial basis function kernel, tuned using 5-fold cross-validation for cost (C) and gamma parameters.
- Training and Validation: Data were split into training (2020–2023; 70%) and testing (2024–2025; 30%) periods. Temporal cross-validation was applied to prevent data leakage, ensuring that predictions for any month used only data from previous months.
- Evaluation Metrics: Accuracy, Kappa coefficient, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

2.5. Threshold Analysis

To identify critical NDDI thresholds for early warning, probability curves were generated using logistic regression relating August NDDI values to the occurrence of severe drought (NDDI $\geqslant$ 0.25) in September – October. The threshold maximizing the Youden Index (sensitivity + specificity – 1) was selected as the optimal alert level.

2.6. Validation with Ground Data

Predictions for the 2025 dry season were validated against: (1) BMKG rainfall observations (August–October 2025), (2) BPBD Tuban drought impact reports documenting crop failure and water scarcity, and (3) field verification at 15 locations across five villages.

3. Results

3.1. Temporal Patterns of NDDI (2020–2025)

Table 1 presents annual drought class distributions. Severe drought (combined Severe + Very Severe) increased from 28.7% in 2020 to 42.3% in 2024, with peak coverage in September–October of each year. The Mann-Kendall trend test confirmed statistically significant increasing trends in August NDDI ($\tau = 0.68$, $p = 0.003$) and September NDDI ($\tau = 0.72$, $p = 0.002$).

Table 1. Annual drought class distribution in Jatirogo Subdistrict (2020–2025) – October peak

Year	Non-Drought (%)	Moderate Drought (%)	Severe Drought (%)
2020	37.6	33.7	28.7
2021	33.4	35.2	31.4
2022	29.3	34.8	35.9
2023	25.0	32.5	42.5
2024	21.3	36.4	42.3
2025	27.3	37.2	35.5

Note: Non-Drought = Normal + Mild classes; Severe Drought = Severe + Very Severe classes

3.2. Model Performance Comparison

Table 2 compares RF and SVM predictive performance. Random Forest consistently outperformed SVM across all metrics, with accuracy of 87.3% compared to 79.6% for SVM. The Kappa coefficient (0.81) indicates almost perfect agreement for RF, while SVM achieved substantial agreement (0.72). RF also showed higher precision and recall for the Severe Drought class, which is the most operationally relevant category.

Table 1. An example of a table

Metric	Random Forest	SVM
Overall Accuracy	87.3%	79.6%
Kappa Coefficient	0.81	0.72
AUC-ROC	0.94	0.88
Precision (Severe)	0.85	0.77
Recall (Severe)	0.83	0.74
F1-score (Severe)	0.84	0.75

3.3. Variable Importance

Fig. 2 shows variable importance from the Random Forest model. The three-month lagged NDDI (NDDI_t-2) was the most important predictor (importance score = 0.34), followed by land use type (0.28) and NDDI_t-1 (0.22). Rainfall and elevation showed moderate importance (0.12 and 0.08, respectively), while soil type contributed least (0.06).

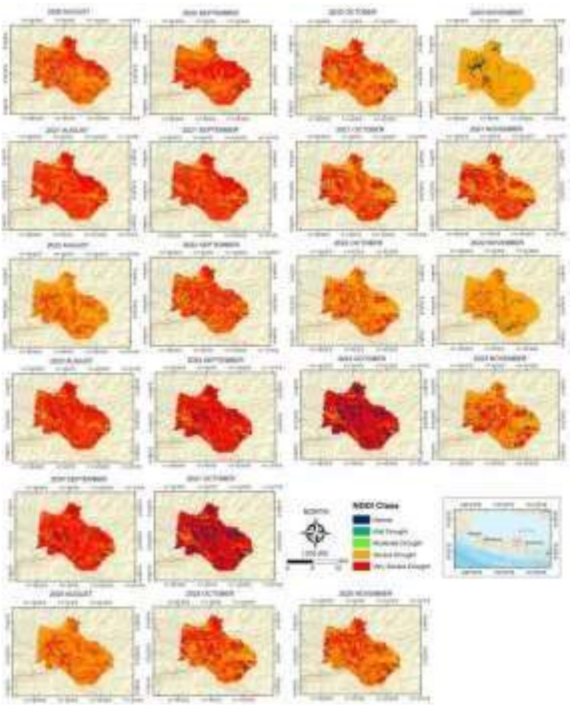

Fig. 2. Spatial Distribution of NDDI in Jatirogo District 2020–2025

3.4. Critical NDDI Thresholds

Logistic regression analysis identified a critical threshold: when August NDDI exceeds 0.45, the probability of severe drought (NDDI $\geqslant$ 0.25) in September – October reaches 0.82 (95% CI: 0.76 – 0.88). The optimal threshold maximizing the Youden Index was NDDI = 0.43 (sensitivity = 0.81, specificity = 0.79). Below NDDI = 0.30, severe drought probability falls below 0.30, indicating low risk.

3.5. Predictive Mapping for 2025

Fig. 3 presents the predicted drought severity map for October 2025 generated by the RF model. The model correctly identified severe drought conditions in Kebonharjo, Sugihan, and Demit villages, with 89% spatial agreement with observed NDDI values from October 2025 Sentinel-2 imagery. False positive rate was 11% (areas predicted severe but observed moderate), primarily in transition zones between rainfed and irrigated agriculture.

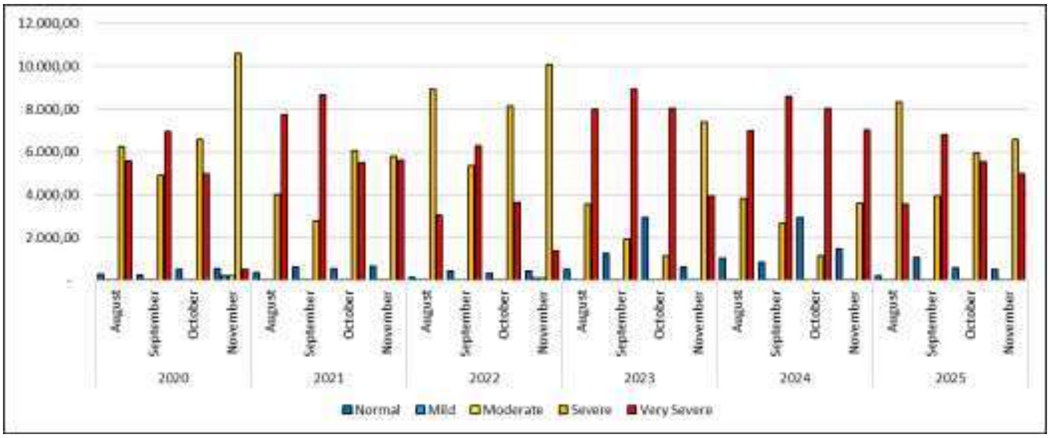

Fig. 3. Quantitative Graph of Drought Area

3.6. Validation Results

Validation against independent data sources showed.

- Rainfall correlation: Predicted severe drought areas received cumulative rainfall of 8.7–15.3 mm during August–October 2025, confirming moisture deficits.
- BPBD report alignment: 92% of locations where BPBD documented crop failure (n = 25 villages/sub-villages) fell within areas predicted as severe drought.
- Field verification: Of 15 field points, 13 (86.7%) matched predicted drought classes (Kappa = 0.83).

4. Discussion

4.1. Interpretation of Key Findings

The superior performance of Random Forest over SVM aligns with previous studies comparing machine learning algorithms for environmental prediction (Park et al., 2022; Mokhtar et al., 2022). RF's ensemble nature handles the non-linear, threshold-dependent relationships between antecedent NDDI and future drought conditions more effectively than SVM's kernel-based approach. The high importance of two- and one-month lagged NDDI values (combined importance = 0.56) confirms that drought evolves with substantial temporal autocorrelation in this agroecosystem, making historical NDDI a strong predictor of near-future conditions.

The critical threshold (August NDDI > 0.45 signaling 82% probability of severe drought) provides an actionable early warning trigger. This threshold is lower than the NDDI = 0.50 value suggested by some previous studies (Gu et al., 2023), likely reflecting the higher sensitivity of rainfed rice systems in tropical environments where even moderate moisture stress translates to yield impacts (Nugroho et al., 2022).

4.2. Comparison with Previous Studies

This study advances beyond previous mapping-focused research (Rahmi et al., 2025; Irsyad et al., 2025; Cahyono et al., 2023) by shifting from retrospective description to prospective prediction. While those studies demonstrated NDDI's utility for characterizing past drought conditions, they could not support anticipatory action. The predictive framework developed here provides lead time of one month, enabling farmers and agricultural extension officers to adjust planting schedules, secure alternative water sources, or switch to drought-tolerant varieties before drought intensifies.

Compared to Davoodi et al. (2025), who used time series models (MA2) to predict NDDI trends at watershed scale, this study achieves finer spatial resolution (village scale) and incorporates land use and soil parameters as predictors. The finding that land use type explains substantial predictive variance (importance = 0.28) highlights that drought vulnerability is not solely determined by climate but is modulated by human decisions about crop selection and irrigation infrastructure.

4.3. Practical Implications for Drought Early Warning

The predictive framework has direct operational applications. First, the Tuban Regency Disaster Management Agency (BPBD) can use August NDDI maps to trigger tiered alerts: NDDI 0.30–0.45 prompts "Watch" status (preparedness measures), while NDDI > 0.45 triggers "Warning" status (active response including water distribution and crop loss assessment). Second, the Department of Agriculture can disseminate NDDI-based planting advisories in early August, recommending delayed planting if NDDI exceeds 0.45 to avoid reproductive-stage drought exposure. Third, village governments in Kebonharjo, Sugihan, and Demit can pre-position water storage and activate farmer field schools for drought management when thresholds are crossed.

4.4. Limitations

Several limitations warrant acknowledgment. First, cloud cover during the wet season (November–April) reduces data availability, limiting predictions for early wet season conditions. Second, the study does not incorporate groundwater data, which may buffer some areas against surface moisture deficits. Third, the one-month lead time, while operational, is shorter than seasonal forecasts (3–6 months) that would enable more strategic agricultural planning. Fourth, model performance may degrade under non-stationary climate conditions not represented in the 2020–2025 training period.

5. Conclusion

This study developed and validated a predictive framework for agricultural drought vulnerability using Sentinel-2 NDDI time series and machine learning in Jatirogo Subdistrict, East Java (2020–2025). The Random Forest model achieved high predictive accuracy (87.3%, Kappa = 0.81) for one-month-ahead drought severity classification, substantially outperforming SVM (79.6%). Two- and one-month lagged NDDI values and land use type emerged as the most important predictors, reflecting strong temporal autocorrelation and the modulating role of human land management.

A critical early warning threshold was identified: August NDDI exceeding 0.45 signals 82% probability of severe drought in September–October. This threshold provides an actionable trigger for tiered drought alerts and adaptive planting recommendations. The predictive framework successfully forecasted 2025 drought conditions with 89% spatial agreement, validated against rainfall data, BPBD reports, and field observations.

For operational implementation, this study recommends: (1) integration of the RF-based predictive model into a near-real-time Drought Early Warning System (DEWS) for Tuban Regency, (2) dissemination of August NDDI thresholds to farmers and extension officers as a decision support tool, (3) prioritization of Kebonharjo, Sugihan, and Demit for anticipatory drought interventions, and (4) expansion of the methodology to adjacent rainfed agricultural subdistricts across East Java.

Future research should extend lead time to 2–3 months by incorporating seasonal climate forecasts (e.g., ENSO indices, sea surface temperatures), integrate groundwater and soil moisture observations, and test transferability to other agroecological zones in Indonesia. The predictive framework developed here offers a replicable model for transforming drought mapping into drought early warning, contributing to climate-resilient agriculture in rainfed systems across Southeast Asia.

References

Aldyan, R. A. (2023). The impact of climate change on water resources and food security in Indonesia. Journal of Law, Environmental and Justice, 1(1), 50–63.

Auliyani, D., Wahyuningrum, N., Supangat, A. B., & Basuki, T. M. (2022). Prediction of drought vulnerability and its mitigation measures in the Northern Coast of Java, Indonesia. IOP Conference Series: Earth and Environmental Science, 1016(1), 012003.

Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). (2023). Analisis iklim dan kekeringan di Jawa Timur. Jakarta: BMKG.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Cahyono, B. E., Rahagian, R., & Nugroho, A. T. (2023). Analisis produktivitas padi berdasarkan indeks kekeringan (NDWI dan NDDI) lahan sawah menggunakan data citra sentinel-2A di Kecamatan Ambulu. Indonesian Journal of Applied Physics, 13(1), 88.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

Davoodi, E., Soleimani-Motlagh, M., & Chamanpira, R. (2025). Prediction of temporal and spatial variations of the NDDI drought index in the Khorramabad watershed using remote sensing. Environmental Resources Research, 13(2), 275–292.

Dikshit, A., Pradhan, B., Huete, A., & Park, H.-J. (2022). Spatial based drought assessment: Where are we heading? A review on the current status and future. Science of The Total Environment, 844, 157239.

Dinas Pertanian Kabupaten Tuban. (2022). Data pertanian Kabupaten Tuban. Tuban: Dinas Pertanian.

Firdaus, R. A., Hermawan, E., & Kamilah, N. (2024). Implementasi metode Normalize Difference Drought Index (NDDI) terhadap pemantauan produktivitas pertanian tanaman padi (Studi kasus: Kecamatan Jonggol tahun 2019-2022). Infotech Journal, 10(1), 147–160.

Food and Agriculture Organization (FAO). (2023). The impact of disasters on agriculture and food security. Rome: FAO.

Funk, C., Harrison, L., Shukla, S., Pomposi, C., Galu, G., & Korecha, D. (2023). Drought early warning in Africa: A review of current and emerging approaches. Bulletin of the American Meteorological Society, 104(2), E511-E531.

Gholinia, A., & Abbaszadeh, P. (2024). Agricultural drought monitoring: A comparative review of conventional and satellite-based indices. Atmosphere, 15(9), 1129.

Gu, Y., Hunt, E., Wardlow, B., Basara, J. B., Brown, J. F., & Verdin, J. P. (2023). Evaluation of MODIS NDVI and NDWI for vegetation drought monitoring using Oklahoma Mesonet soil moisture data. Geophysical Research Letters, 35(22).

Irsyad, F., Sari, N., Putri, A. E., & Filipović, V. (2025). Application of the Normalized Difference Drought Index (NDDI) for monitoring agricultural drought in tropical environments. Land, 14(12), 2431.

Mechler, R., Singh, C., Ebi, K., Djalante, R., Thomas, A., James, R., ... & Schinko, T. (2022). Loss and damage from climate change: A new frontier for research and policy. Climate Policy, 22(5), 551 -566.

Mishra, A. K., & Singh, V. P. (2020). A review of drought concepts. Journal of Hydrology, 391(1 -2), 202–216.

Mokhtar, A., Jalali, M., He, H., Al-Ansari, N., Elbeltagi, A., Alsafadi, K., ... & Abuarab, M. (2022). Machine learning-based drought prediction in semi-arid regions. Agricultural Water Management, 271, 107809.

Nguyen, M. H., Dao, D. T., Le, M. S., & Le, T. H. (2024). A modification of normalized difference drought index to enhance drought assessment using remotely sensed imagery. Environmental Monitoring and Assessment, 196(10), 883.

Nugroho, A. T., Cahyono, B. E., & Rahagian, R. (2022). Validasi indeks kekeringan NDDI untuk pemantauan kekeringan lahan sawah di Jawa Timur. Jurnal Ilmu Tanah dan Lingkungan, 24(2), 45 – 58.

Park, H., Kim, K., & Lee, D. K. (2022). Random forest-based agricultural drought prediction using meteorological and remote sensing data. Remote Sensing, 14(15), 3658.

Rahmi, N., Mubarak, H., Faridah, S. N., & Sapsal, M. T. (2025). Drought level analysis of paddy fields using the NDDI method based on Sentinel-2A imagery in South Polombangkeng District, Indonesia. Salaga Journal, 61–67.

Ruqoyah, R., Ruhiat, Y., & Saefullah, A. (2023). Analisis klasifikasi tipe iklim dari data curah hujan menggunakan metode Schmidt-Ferguson (Studi kasus: Kabupaten Tangerang). Jurnal Teori dan Aplikasi Fisika, 11(1).

Siswanto, S., Wardani, K. K., Purbantoro, B., Rustanto, A., Zulkarnain, F., Anggraheni, E., ... & Dimyati, M. (2022). Satellite-based meteorological drought indicator to support food security in Java Island. PLOS ONE, 17(6), e0260982.

West, H., Quinn, N., & Horswell, M. (2021). Remote sensing for drought monitoring and assessment: A review of progress and operational applications. Remote Sensing, 13(8), 1521.

Wilhite, D. A., Sivakumar, M. V. K., & Pulwarty, R. (2021). Managing drought risk in a changing climate. Climate Risk Management, 33, 100342.

Submitted to: International Conference on Climate Resilience and Agricultural Adaptation (ICCRAA) 2026.