Lite Learning: Efficient Crop Classification in Tanzania Using Feature Extraction with Machine Learning & Crowd Sourcing

Michael L. Mann, Lisa Colson, Rory Nealon, Ryan Engstrom, Stellamaris Nakacwa

***Abstract*—This study introduces a novel approach to traditional machine learning methodology for crop type classification in Tanzania, by integrating crowdsourced data with time-series features extracted from Sentinel-2 satellite imagery. Leveraging the YouthMappers network, we collected ground validation data on various crops, including challenging types such as cassava, millet, sunflower, sorghum, and cotton across a range of agricultural areas. Traditional machine learning algorithms, augmented with carefully engineered time-series features, were employed to map the different crop classes. Our approach achieved high classification accuracy, evidenced by a Cohen's Kappa score of 0.82 and an F1-micro score of 0.85. The model often match or outperform broadly used land cover models which simply classify 'agriculture' without specifying crop types. By interpreting feature importance using SHAP values, we identified key time-series features driving the model's performance, enhancing both interpretability and reliability. Our findings demonstrate that traditional machine learning techniques, combined with computationally efficient feature extraction methods, offer a practical and effective "lite learning" approach for mapping crop types in data-scarce environments. This methodology facilitates accurate crop type classification using a low-cost, resource-limited approach that contributes valuable insights for sustainable agricultural practices and informed policy-making, ultimately impacting food security and land management in resource-limited contexts, such as sub-Saharan Africa.**

Michael L. Mann and Ryan Engstrom are with The George Washington University, Washington DC 20052 (e-mail: mmann1123@gmail.com).

Lisa Colson is with the USDA Foreign Agricultural Service, Washington DC 20250.

Rory Nealon is with USAID GeoCenter, Washington DC 20523.

Stellamaris Nakacwa is with YouthMappers, Texas Tech University, Lubbock TX 79409.

***Index Terms*—Remote sensing, machine learning, crop classification, time-series analysis, crowdsourcing, Sentinel-2, feature engineering.**

I. Introduction

A. Background and Context

The free access to remotely sensed data, such as imagery from satellites (e.g. Sentinel-2, Landsat), has allowed for crop type classification in developing countries. By leveraging the power of advanced imaging technologies combined with machine learning algorithms, researchers and practitioners can now identify and map different crop types over large geographic areas at no or low cost [1]. This has the potential to improve food security, land use planning, and agricultural policy in regions where ground-based data collection is limited or non-existent [2]–[4].

In recent years, machine learning approaches have emerged as powerful tools for crop type classification using remotely sensed data. Specifically, methods based on machine learning algorithms have gained recognition for their effectiveness in matching valuable spectral information from satellite imagery to observations of crop type for particular locations. Machine learning algorithms, including decision trees, random forests, support vector machines (SVM), and k-nearest neighbors (KNN), have been successfully used to classify imagery into unique agricultural types [4]–[6]. These algorithms leverage the rich spectral information captured by satellite sensors, allowing them to identify distinctive patterns associated with different crop types.

By training on large labeled datasets where ground-validation information on crop types is linked to corresponding image pixels, these models can effectively learn the relationships between the spectral characteristics of crops and their respective classes [2].

The strength of traditional machine learning approaches lies in their ability to exploit both the spectral and time-series patterns within the remotely sensed data. Traditional machine learning approaches offer advantages in terms of interpretability and computational efficiency compared to deep learning architectures. They provide insight into the decision-making process and can be more readily understood and explained by domain experts. Additionally, these methods are generally less computationally demanding and require less training data, making them suitable for applications with limited computational resources [7]–[11].

Traditional machine learning algorithms require the extraction of variables (e.g. max EVI, mean blue band) that can help distinguish different plant or crop types [5]. The development of salient time-series features to capture phenological differences between locations from remotely sensed images remains a challenge. These features are typically derived from the spectral bands (e.g. red edge, NIR) of the satellite imagery or indexes, such as the enhanced vegetation index (EVI), and basic time series statistics (e.g. mean, max, minimum, slope) for the growing season [12]. Meanwhile a broader set of time series statistics from bands or indexes may be more relevant for a number of applications. For instance the skewness of EVI might help distinguish crops that green-up earlier vs later in the season, measures of the numbers of peaks in EVI might help differentiate intercropping or multiple plantings in a season [5]. However, the selection and extraction of these features can be time-consuming and labor-intensive, requiring domain expertise and manual intervention.

In contrast, deep learning methods have dominated the most recent literature [7], [9]. These methods include both recurrent neural networks (RNN) and convolutional neural networks (CNN). Recurrent Neural Networks (RNNs) are a class of neural networks that are particularly powerful for modeling sequential data such as time series, speech, text, and audio. The fundamental feature of RNNs is their ability to maintain a 'memory' of previous inputs by using their internal state (hidden layers), which allows them to exhibit dynamic temporal behavior. RNNs and its variants allow the integration of time-series imagery, significantly improving crop type classification outcomes especially in data rich environments [9], [13]. Deep learning approaches however typically require much larger sets of training data, may be more prone to overfitting especially with small sample sizes, have significant limitations to interpretability, and require expensive compute [7]–[11]. Although recent efforts have closed the gap e.g. [14], the lack of readily available and reliable ground truth data or benchmark datasets for training, as discussed earlier, may limit the applicability of deep learning for a variety of tasks including crop classification and make researchers more reliant of less reliable techniques like transfer learning or zero-shot or low shot methods [10], [11], [15]. Moreover, training data for extreme events, like crop losses, disease, and lodging are largely non-existant. Interpretability is also a salient weakness as interpretation of models allows us to gain scientific insight and assess trustworthiness and fairness in so far as outputs affect policy decisions.

An alternative approach turns back the clock on deep learning approaches. For instance CNN classifiers, through the exertion of tremendous effort of GPUs, can apply and learn from thousands of filters or convolutions that help detect distinct features like edges, textures or patterns. It is however possible to apply a more limited yet salient set of filters like Fourier Transforms, Differential Morphological Profiles [16], Line Support Regions or Structural Feature Sets [17] amongst others, to images and then use these as features in more traditional machine learning approaches [15], [18]–[21]. This approach may be particularly useful in data-scarce environments, requiring less training data and potentially offering more efficient results in low-information settings. The same approach has been taken for time series analysis, where instead of learning patterns through a RNNs memory, we can apply a more limited but potentially salient series of time series filters. Measures of trends, descriptions of distributions, or measures of change and complexity might adequately describe time series properties for regression and classification tasks [22], [23]. This time series filter approach, developed for this paper, can also be applied on a pixel-by-pixel basis to satellite image bands or index values [24].

Field-collected data provides the necessary validation and calibration for remote sensing-based models. It serves as the benchmark against which the model's predictions are evaluated and refined. Ground validation data collected through field visits, observation, and interactions with local farmers offers essential insights into the specific crop types present in the study area. Validating and training models with accurate ground reference information allows for the spectral patterns captured by remote sensing data to be correctly associated with the corresponding crop. By combining the spectral information from satellite imagery with ground validation data, researchers can develop robust models that effectively differentiate between different crop types based on their unique spectral signatures and temporal patterns.

The collection of field observations and ground validation data is a critical input for the development of models to classify crop types [6], [11]. However, obtaining accurate and timely ground validation data can be challenging in developing countries due to limited resources, infrastructure, and local capacity [5], [6]. In many cases, researchers rely on crowdsourced data from volunteers or citizen scientists to supplement or validate ground truth data collected through traditional methods. Projects like [14] point to the paucity of multi-class crop type datasets globally. This is a significant gap in the field of crop type classification, as the availability of high-quality training data is essential for the development of accurate and reliable machine learning models [8].

In this study we aim to address two critical challenges in the field of crop type classification: the lack of in-season multi-class crop type datasets, and the need for new methods to obtain high accuracy crop type predictions from limited amounts of training data.

We propose a novel approach that combines crowdsourced data with a new automated approach to extracting time-series features from satellite imagery. We apply this new approach to classify crop types in Northern Tanzania. By leveraging the power of crowdsourcing and remote sensing technologies, we aim to develop a robust and scalable solution for crop type classification that can be adapted to other regions and contexts with a minimal or no cost.

II. Data & Methods

Data for this study were collected from multiple sources, including satellite imagery, and crowdsourced ground truth observations. The section below describes the input data and methods used throughout the paper.

A. *Study Area*

The study was conducted in 50 wards within three major districts of Arusha, Dodoma and Mwanza in Tanzania as seen in Figure 1. Tanzania, a country in East Africa, is known for its diverse agricultural landscape. The region is characterized by a mix of smallholder farms, commercial plantations, and natural vegetation, making it an ideal yet challenging location for studying crop type classification. Our choice of these three districts was driven by the distinct variation in the major crop types that possibly dominated in each district,among oil seeds, grains and commercial crops such as cotton.

B. *Crowd Sourced Data Collection*

Crop type data collection was designed and executed by YouthMappers through a crowdsourced GIS approach. The method was designed in 3 steps where: 1) Development of and training all intended student participants. 2) Data collection using KoboToolbox hosting a well developed data model. The exercise lasted 14 days with 7 days of iterative pilot testing on different farms, crops and landscapes. Finally the last step, 3) was the data

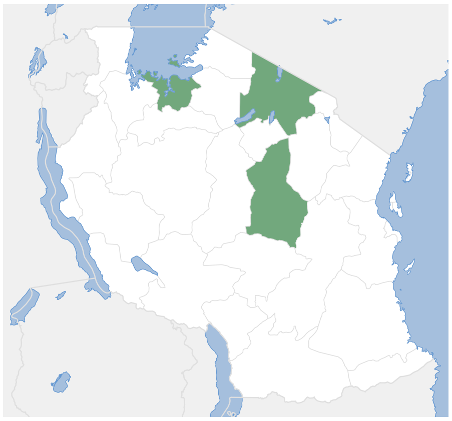

Fig. 1: Study area map
Districts in Northern Tanzania where field visits were carried out (green)

review and cleaning phase to generate a sample for training.

Additional training data was collected utilizing high resolution imagery from Google Earth. These data were used to supplement the crowdsourced data and improve the model's ability to distinguish between crops and more common land cover types like forests, urban areas, and water. The final cleaned dataset includes 1,400 crop type observations of rice, maize, cassava, sunflower, sorghum, cotton, and millet; plus 386 other observations of land cover classes including water, tidal areas, forest, shrub and urban.

1) Data Collection Methods: To ensure the success of our project, we focused heavily on the design of our data collection methods. These methods were carefully integrated, taking into account: the crop calendar, information on the different stages of crop development, the distances between crop fields, the tools used, and data quality assurance.

Young crops exhibit significant differences compared to mature crops in terms of color, density, and phenological development. Variations in the crop cycle across different fields could lead to heteroscedasticity in the spectral reflectance measurements used for machine learning (ML) training, thereby affecting the precision and accuracy of the model. By targeting the period of April through

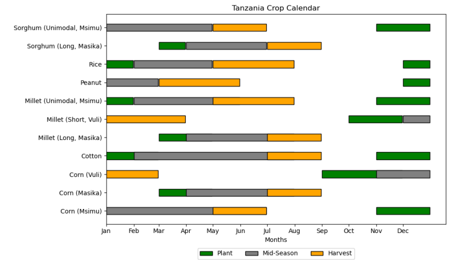

Fig. 2: Tanzania Crop Calendar

May we aimed to capture crops late in the growing season and yet before harvest as seen in the crop calendar in Figure 2 below.

Source: [25]

USDA's Foreign Agricultural Service compiles information on planting and harvest windows for grain, oilseed, and cotton crops as an important tool to support crop condition assessments with satellite imagery. Tanzania's crop planting seasons are shaped by its bimodal and unimodal rainfall patterns, which vary by region. In the north and northeast, bimodal areas experience the short rains (Vuli) from late-October to mid-January, during which crops like maize, beans, and vegetables are planted in October and November, and the long rains (Masika) from March to May, supporting crops like maize, rice, sorghum, and cassava, typically planted in February and March. In the central, southern, and western regions with unimodal rainfall, there is a single rainy season from November to April, when crops such as cotton, maize, millet, rice, and sunflower are planted in November and December. This diversity in rainfall patterns allows for a wide variety of crops suited to the local climate and seasonal conditions.

The data collection took place between late April and May as shown in figure 2 to align with mid-season for many crops. YouthMappers were advised to focus on a set of target crops, ones known to be present in the region and at appropriate crop growth stages. Before embarking on data collection, discussions covered several factors to consider in selecting field collection sites. Factors included field size to establish a minimum detectable by the satel-

0.84 across a diverse and multi-class dataset. The model accurately classified challenging crops such as cassava, millet, sorghum, and cotton. The integration of crowdsourced data and time-series features provided valuable insights into the temporal dynamics of crop growth, enhancing the model's accuracy and reliability. Notably, our model—although trained specifically on Tanzanian data—outperforms broadly used land cover models that perform the simpler task of classifying 'agriculture' without specifying the crop type. This highlights the need for better and more frequent crop type classification data.

By interpreting feature importance using SHAP values, we gained a deeper understanding of the model's behavior and the key predictors driving its predictions. Identifying the most influential features across different land cover types allowed us to refine the feature selection process, ensuring that the selected features were both consistently influential and impactful under specific conditions.

In conclusion, our study underscores the viability and effectiveness of traditional machine learning approaches augmented with carefully engineered time-series features for crop type classification in data-scarce environments. By "turning back the clock" on deep learning, we demonstrate that applying a limited yet salient set of filters—such as measures of trends, distribution descriptions, and complexity metrics—can capture essential temporal dynamics without the extensive data requirements of deep learning models. This methodology not only achieved high classification accuracy but also enhanced interpretability and computational efficiency.

Our findings highlight that traditional machine learning techniques, combined with advanced yet computationally efficient feature extraction methods, offer a practical and effective alternative to deep learning, particularly in low-information settings prevalent in developing regions. This approach facilitates accurate crop type classification and contributes valuable insights for sustainable agricultural practices and informed policy-making, ultimately impacting food security and land management in resource-limited contexts.

References

[1] J. Hersh, R. Engstrom, and M. Mann, "Open data for algorithms: mapping poverty in belize using open satellite derived features and machine learning," *Information Technology for Development*, vol. 27, no. 2, pp. 263–292, 2021.

[2] A. Bégué, D. Arvor, B. Bellon, J. Betbeder, D. De Abelleyra, R. P. D. Ferraz, V. Lebourgeois, C. Lelong, M. Simões, and S. R. Verón, "Remote sensing and cropping practices: A review," *Remote Sensing*, vol. 10, no. 1, 2018. [Online]. Available: https://www.mdpi.com/2072-4292/10/1/99

[3] H. Li, X.-P. Song, M. C. Hansen, I. Becker-Reshef, B. Adusei, J. Pickering, L. Wang, L. Wang, Z. Lin, V. Zalles *et al.*, "Development of a 10-m resolution maize and soybean map over china: Matching satellite-based crop classification with sample-based area estimation," *Remote Sensing of Environment*, vol. 294, p. 113623, 2023.

[4] E. S. Ibrahim, P. Rufin, L. Nill, B. Kamali, C. Nendel, and P. Hostert, "Mapping crop types and cropping systems in nigeria with sentinel-2 imagery," *Remote sensing*, vol. 13, no. 17, p. 3523, 2021.

[5] A. Bégué, D. Arvor, B. Bellon, J. Betbeder, D. De Abelleyra, R. PD Ferraz, V. Lebourgeois, C. Lelong, M. Simões, and S. R. Verón, "Remote sensing and cropping practices: A review," *Remote Sensing*, vol. 10, no. 1, p. 99, 2018.

[6] J. Delince, G. Lemoine, P. Defourny, J. Gallego, A. Davidson, S. Ray, O. Rojas, J. Latham, and F. Achard, "Handbook on remote sensing for agricultural statistics," *GSARS: Rome, Italy*, 2017.

[7] A. Höhl, I. Obadic, M.-Á. Fernández-Torres, D. Oliveira, and X. X. Zhu, "Recent trends challenges and limitations of explainable ai in remote sensing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 8199–8205.

[8] A. E. Maxwell, T. A. Warner, and L. A. Guillén, "Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—part 2: Recommendations and best practices," *Remote Sensing*, vol. 13, no. 13, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/13/2591

[9] I. Teixeira, R. Morais, J. J. Sousa, and A. Cunha, "Deep learning models for the classification of crops in aerial imagery: A review," *Agriculture*, vol. 13, no. 5, 2023. [Online]. Available: https://www.mdpi.com/2077-0472/13/5/965

[10] Y. Li, X. Li, Y. Zhang, D. Peng, and L. Bruzzone, "Cost-efficient information extraction from massive remote sensing data: When weakly supervised deep learning meets remote sensing big data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 120, p. 103345, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S156984322300167X

[11] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, pp. 166–177, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271619301108

[12] D. C. Morton, R. S. DeFries, Y. E. Shimabukuro, L. O. Anderson, E. Arai, F. del Bon Espirito-Santo, R. Freitas, and J. Morisette, "Cropland expansion changes deforestation dynamics in the southern brazilian amazon," *Proceedings of the National Academy of Sciences*, vol. 103, no. 39, pp. 14 637–14 641, 2006.

[13] G. Camps-Valls, D. Tuia, X. X. Zhu, and M. Reichstein, *Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science and geosciences*. John Wiley & Sons, 2021.

[14] G. Tseng, I. Zvonkov, C. L. Nakalembe, and H. Kerner, "Cropharvest: A global dataset for crop-type classification," in *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021. [Online]. Available: https://openreview.net/forum?id=JtjzUXPEaCu

[15] M. Owusu, A. Nair, A. Jafari, D. Thomson, M. Kuffer, and R. Engstrom, "Towards a scalable and transferable approach to map deprived areas using sentinel-2 images and machine learning," *Computers, Environment and Urban Systems*, vol. 109, p. 102075, 2024.

[16] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 309–320, 2001.

[17] X. Huang, L. Zhang, and P. Li, "Classification and extraction of spatial features in urban areas using high-resolution multispectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, pp. 260–264, 2007.

[18] J. Graesser, A. Cheriyadat, R. R. Vatsavai, V. Chandola, J. Long, and E. Bright, "Image based characterization of formal and informal neighborhoods in an urban landscape," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 4, pp. 1164–1176, 2012.

[19] R. Engstrom, J. Hersh, and D. Newhouse, "Poverty from space: Using high resolution satellite imagery for estimating economic well-being," *The World Bank Economic Review*, vol. 36, no. 2, pp. 382–412, 2022.

[20] S. Chao, R. Engstrom, M. Mann, and A. Bedada, "Evaluating the ability to use contextual features derived from multi-scale satellite imagery to map spatial patterns of urban attributes and population distributions," *Remote Sensing*, vol. 13, no. 19, p. 3962, 2021.

[21] M. Owusu, R. Engstrom, D. Thomson, M. Kuffer, and M. L. Mann, "Mapping deprived urban areas using open geospatial data and machine learning in africa," *Urban Science*, vol. 7, no. 4, 2023. [Online]. Available: https://www.mdpi.com/2413-8851/7/4/116

[22] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package)," *Neurocomputing*, vol. 307, pp. 72–77, 2018.

[23] Z. Yang, I. A. Abbasi, E. E. Mustafa, S. Ali, and M. Zhang, "An anomaly detection algorithm selection service for iot stream data based on tsfresh tool and genetic algorithm," *Security and Communication Networks*, vol. 2021, no. 1, p. 6677027, 2021.

[24] Mann, Michael L., "xr_fresh: Python Package for Feature Extraction from Raster Data Time Series," 2024, xr_fresh is designed to quickly generate a broad set of temporal features from gridded raster data time series. This package works only with single band images, not multiband (e.g., RGB) formats. It leverages Python and xarray for analyzing time-series characteristics from raster data as generated by geowombat. Capabilities include feature extraction, dimension reduction, and application of machine learning techniques for geospatial data. Input data: historical raster data (e.g., Monthly temperature data from 2000-2018). Output: rasters with features as columns or raster file with features as bands. [Online]. Available: https://doi.org/10.5281/zenodo.12701466

[25] FAS, "Tanzania production," https://ipad.fas.usda.gov/countrysummary/default.aspx?id=TZ. [Online]. Available: https://ipad.fas.usda.gov/countrysummary/default.aspx?id=TZ

[26] Google, "Ee.image.rgbtohsv google earth engine," *Google Earth Engine*, 2025. [Online]. Available: https://developers.google.com/earth-engine/apidocs/ee-image-rgbtohsv

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[28] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[29] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4765–4774. [Online]. Available: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[30] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, 2017.

[31] H. Kerner, C. Nakalembe, A. Yang, I. Zvonkov, R. McWeeny, G. Tseng, and I. Becker-Reshef, "How accurate are existing land cover maps for agriculture in sub-saharan africa?" *Scientific Data*, vol. 11, no. 1, p. 486, 2024.

APPENDIX

A5. ACKNOWLEDGMENTS

The United States Agency for International Development generously supports this program through a grant from the USAID GeoCenter under Award # AID-OAA-G-15-00007 and Cooperative Agreement Number: 7200AA18CA00015.

A6. TIME SERIES FEATURES DESCRIPTION

The following table provides a comprehensive list of the time series features extracted from the satellite imagery using the `xr_fresh` module. These features capture the temporal dynamics of crop growth and development, providing valuable information on the phenological patterns of different crops. The computed metrics encompass a wide range of statistical measures, changes over time, and distribution-based metrics, offering a detailed analysis of the temporal patterns in the study area.

Statistic	Description	Equation
Absolute energy	sum over the squared values	$E = \sum_{i=1}^{n} x_i^2$
Absolute Sum of Changes	sum over the absolute value of consecutive changes in the series	$\sum_{i=1}^{n-1} \mid x_{i+1} - x_i \mid$
Autocorrelation (1 & 2 month lag)	Correlation between the time series and its lagged values	$\frac{1}{(n-l)\sigma^2} \sum_{t=1}^{n-l} (X_t - \mu)(X_{t+l} - \mu)$
Count Above Mean	Number of values above the mean	$N_{\text{above}} = \sum_{i=1}^{n} (x_i > \bar{x})$
Count Below Mean	Number of values below the mean	$N_{\text{below}} = \sum_{i=1}^{n} (x_i < \bar{x})$
Day of Year of Maximum Value	Day of the year when the maximum value occurs in series	—
Day of Year of Minimum Value	Day of the year when the minimum value occurs in series	—
Kurtosis	Measure of the tailedness of the time series distribution	$G_2 = \frac{\mu_4}{\sigma^4} - 3$
Linear Time Trend	Linear trend coefficient estimated over the entire time series	$b = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(t_i - \bar{t})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$
Longest Strike Above Mean	Longest consecutive sequence of values above the mean	—
Longest Strike Below Mean	Longest consecutive sequence of values below the mean	—
Maximum	Maximum value of the time series	x_{max}
Mean	Mean value of the time series	$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$
Mean Absolute Change	Mean of absolute differences between consecutive values	$\frac{1}{n-1} \sum_{i=1}^{n-1} \lvert x_{i+1} - x_i \rvert$
Mean Change	Mean of the differences between consecutive values	$\frac{1}{n-1} \sum_{i=1}^{n-1} x_{i+1} - x_i$
Mean Second Derivative Central	measure of acceleration of changes in a time series data	$\frac{1}{2(n-2)} \sum_{i=1}^{n-1} \frac{1}{2} (x_{i+2} - 2 \cdot x_{i+1} + x_i)$
Median	Median value of the time series	$\tilde{x}$
Minimum	Minimum value of the time series	x_{min}
Quantile (q = 0.05, 0.95)	Values representing the specified quantiles (5th and 95th percentiles)	$Q_{0.05}, Q_{0.95}$
Ratio Beyond r Sigma (r=1,2,3)	Proportion of values beyond r standard deviations from the mean	$P_r = \frac{1}{n} \sum_{i=1}^{n} (\lvert x_i - \bar{x} \rvert > r\sigma_x)$

Statistic	Description	Equation
Skewness	Measure of the asymmetry of the time series distribution	$\frac{n}{(n-1)(n-2)} \sum \left(\frac{X_i - \overline{X}}{s} \right)^3$
Standard Deviation	Standard deviation of the time series	$\sqrt{\frac{1}{N} \sum_{i=1}^{n} (x_i - \bar{x})^2}$
Sum Values	Sum of all values in the time series	$S = \sum_{i=1}^{n} x_i$
Symmetry Looking	Measures the similarity of the time series when flipped horizontally	$\lvert x_{\text{mean}} - x_{\text{median}} \rvert < r * (x_{\text{max}} - x_{\text{min}})$
Time Series Complexity (CID CE)	measure of number of peaks and valleys	$\sqrt{\sum_{i=1}^{n-1} (x_i - x_{i-1})^2}$
Variance	Variance of the time series	$\sigma^2 = \frac{1}{N} \sum_{i=1}^{n} (x_i - \bar{x})^2$
Variance Larger than Standard Deviation	check if variance is larger than standard deviation	$\sigma^2 > 1$