Predicting Wildfires in Tanzania Using a Machine Learning Approach

Konstantin Engelmayer
Universitaet Marburg

Word count: approximately 3,677 words.

September 29, 2024

Abstract

Wildfires are a natural, recurring phenomenon in Tanzania's ecosystem. However, due to climate change and human expansion, they increasingly pose a destructive risk to the country's environment and population. This study develops a machine learning model to predict wildfire susceptibility across Tanzania, using data on human activity, as well as climatic and environmental factors. Land cover type emerged as the most influential predictor, highlighting the critical role of vegetation and human activities in wildfire dynamics. The model achieved an overall accuracy of 85.2%, performing well during dry season but less effectively during wet season due to limited climatic variability in the training data. The findings underscore the importance of localized fire risk education and the integration of predictive modeling into fire management strategies. Limitations include the need for higher-resolution climate data and a more balanced representation of seasonal climate conditions. Future research should focus on refining the model to improve responsiveness to climatic variations, enhancing wildfire prediction accuracy and improving wildfire management practices in Tanzania.

Contents

1 Introduction

Wildfires are a natural, recurring phenomenon in Tanzania, shaping ecosystems like the Serengeti, where periodic fires are essential for maintaining grasslands that support wildlife [1, 2]. Such fires clear old vegetation and promote new growth [3, 4]. However, in many forested areas not adapted to fire, wildfires cause significant damage. This risk is exacerbated by the intensifying impacts of climate change [5] as rising temperatures and prolonged droughts provide favorable conditions for wildfires to ignite and spread [6].

Moreover, human expansion adds complexity to fire management. As Tanzania's population grows [7], more land is used for housing, agriculture, and other activities, amplifying the conflict between natural fire regimes and human land use. Uncontrolled wildfires can result in economic and social losses, especially when they affect farmlands, homes, and infrastructure [8].

Frequently, wildfires are accidentally ignited by activities such as farming, honey harvesting, or charcoal burning [9, 10]. In this context, wildfire susceptibility mapping is crucial for containment efforts. Wildfires are not inherently harmful but can be destructive where human activity overlaps with fire-prone ecosystems. The identification of high-risk areas enables local communities and authorities to take precautions, improve early warning systems, and manage fire hazards [11]. Knowing where and when fires are likely to occur enhances decision-making and fire management, reducing destructive impacts on people and the environment [11].

Historically, fire detection has relied on watchtower observers scanning for smoke and flames, but modern technology has replaced this method [12]. Satellite data now provides near-real-time fire monitoring through systems like EUMETSAT, offering fire indices that detect hotspots and assess fire risks, improving detection and response accuracy [12].

Machine learning has recently enhanced wildfire prediction by leveraging large datasets from satellites and weather stations [13]. The resultant models assess fire risks with greater precision, improving the accuracy of wildfire susceptibility mapping [13]. This technology is vital for navigating and reducing the conflict between natural wildfire regimes and human land use.

There are many fire indices and fire risk maps calculated on the global scale, some of which incorporate machine learning. Despite the significant variation in fire dynamics across countries and ecosystems [14], these models are not typically adapted to specific national contexts like Tanzania. The resultant one-size-fits-all approach may overlook critical local factors influencing fire behavior. This study aims to fill this gap by developing a machine learning model that generates wildfire susceptibility maps specifically for Tanzania, utilizing climate data and other environmental factors. By focusing on the national level, this study aims to identify the main drivers for wild fires in Tanzania and enhance preparedness for fires. This would allow an improved mitigation of risks and more effective responses, in turn reducing potential destructive impacts on livelihoods and the environment.

2 Methods

2.1 Research Area

Tanzania, located in East Africa, is bordered by Kenya, Uganda, Rwanda, Burundi, the Democratic Republic of Congo, Zambia, Malawi, Mozambique, and the Indian Ocean. It covers approximately 945,000 square kilometers and features diverse landscapes, including coastal plains, highland plateaus, volcanic mountains, savannahs, and extensive Miombo woodlands found primarily in the western and southern regions of the country (figure 2).

The climate is tropical with regional variations that strongly influence wildfire patterns. Coastal areas are hot and mostly humid, while more moderate temperatures prevail in the inland plateau. Rainfall is bi-modal,

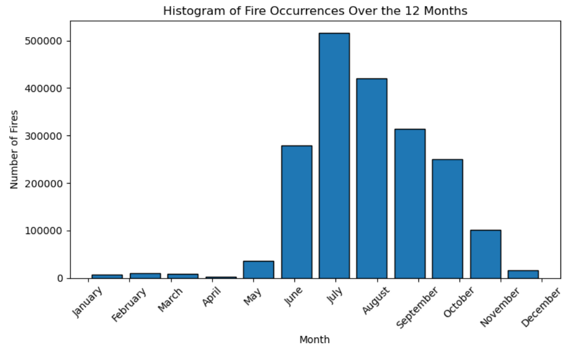


Figure 1: Fire occurrence in Tanzania from 2000 to 2023 for each month showing the fire season

with the wet season lasting from November to May[15].

Wildfires mostly occur during the dry season, from June to October, when vegetation is dry and more flammable (figure 1). Both natural and human factors drive fire occurrence, often in conjunction with one another. Human activities like slash-and-burn agriculture, charcoal production, honey harvesting, and land clearing frequently cause the first ignition of fires.

2.2 Fire and Non-Fire Data

For this study, active fire detections were obtained from the MODIS Collection 6 (MCD14ML) dataset [16], documenting over 1 million fire events in Tanzania between 2000 and 2023. To focus on relevant fires, only those with confidence levels above 80% and classified as 'presumed vegetation fire' were selected, reducing the dataset to 320,000 records. To further improve the reliability of the dataset, annual land cover data from the Copernicus Climate Change Service [17] was also used, again ensuring that only vegetation fires were included. Finally, the fire data was both temporally and spatially aggregated to eliminate duplicate events, leaving 250,000 data points for the study.

The number of non-fire pixels was proportionally selected for each month and year based on the number of fire pixels in the corresponding period. The selection also ensured that no non-fire pixel matched the coordinates of any fire pixel. This is a common, widely used approach in previous studies [18, 19].

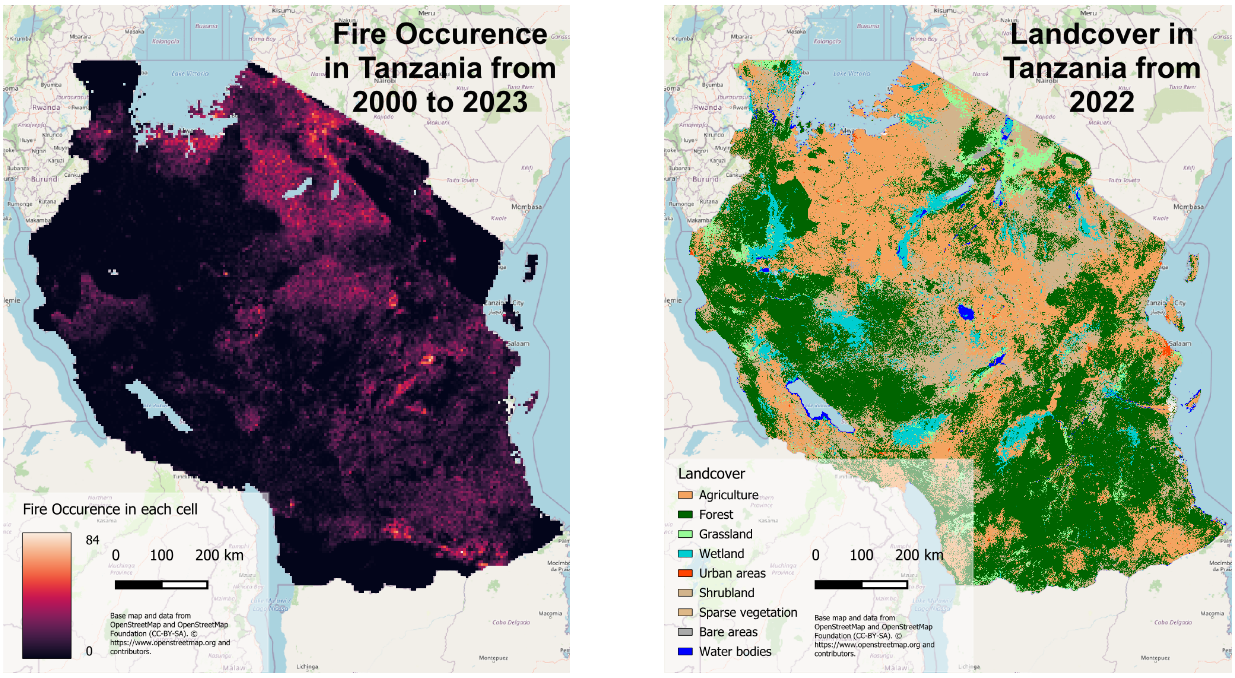
Due to computational limitations, the dataset was then randomly sampled down to 200,000 fire pixels and 200,000 non-fire pixels. This approach ensured a sufficient number of data points for both training and testing the model while significantly reducing the calculation time, allowing for more efficient processing without sacrificing model reliability.

2.3 Predictor Selection

In this study, various datasets were selected to support wildfire prediction in Tanzania based on their relevance to fire susceptibility. Land cover data from the Copernicus Climate Change Service [17], spanning from 2000 to 2022, were incorporated as vegetation type and land use directly affect fuel availability and fire spread potential [20]. Monthly aggregated climate data, including temperature and precipitation, sourced from ERA5-Land [15], were included because temperature influences the likelihood of ignition, while precipitation impacts fuel moisture, both of which are critical factors in wildfire risk. The ERA5-Land data also included the leaf area index (LAI), which gives insights into vegetation density and fuel availability - two key components controlling biomass levels. This further aids in assessing areas with higher fire risk due to greater biomass.

Topographic variables such as the topographic wetness index (TWI), heat load index (HLI), slope, and aspect were derived from the SRTM digital elevation model (DEM) [21]. Slope and aspect influence fire spread [22], while TWI and HLI provide information on water availability and heat exposure [23], both of which affect fire behavior.

Calculated Euclidean distance to streets, based on OpenStreetMap data [24], was also incorporated, as human infrastructure and related human activity is often associated with fire ignition. River data from OpenStreetMap [24] was also included to calculate the distance



(a) (b)

Figure 2: (a) Fire occurrence in Tanzania from 2000 to 2023 in each pixel with a pixel size of 5500x5500m (b) land cover in Tanzania 2022 by Copernicus Climate Change Service

to the nearest river, as proximity to rivers can elevate soil and air moisture levels, potentially reducing fire risk [25].

Vegetation dynamics were captured using MODIS Terra Vegetation Indices (MOD13A3.061) [26], which provided monthly Normalized Difference Vegetation Index (NDVI) data. This dataset was included as vegetation density and stress levels are closely linked to fire susceptibility. Lastly, fire danger indices were retrieved from the Copernicus Emergency Management Service [27] to provide an assessment of environmental conditions conducive to fire risk. These predictors collectively address key factors such as fuel availability, ignition likelihood, and conditions for fire spread, all of which are essential for accurate wildfire susceptibility mapping. (Appendix Table 1)

Many of the predictor variables are highly correlated. To address this, the Relief-F method was applied for an initial selection of the 15 most important variables. Relief-F, developed by Kira and Rendell [28] and later refined by Kononenko [29], is widely recognized for feature selection [30]. This method evaluates the relevance of features by comparing the feature values of each instance with its nearest neighbors from both the same class (referred to as "hits") and different classes (referred to as "misses"). Features that can better distinguish between instances of different classes receive higher weights, while those that do not contribute much to the classification task receive lower weights [29]. This helps identify the most influential features, even in the presence of noisy or correlated data [18]. After the initial feature selection with Relief-F, a forward feature selection (FFS) was applied to further refine the variable selection, ensuring that only the most relevant features were included in the final model. Due to computational constraints, feature selection was performed on a 10% subset of the data (40,000 data points).

2.4 Model Training, Testing and Evaluation

The data was split at 2016, with the training set covering 2000 to 2015 and the test set from 2016 to 2022. This split was chosen to ensure that roughly 75% of the data was used for training and 25% for testing.

With the selected features, a grid-search-based hyperparameter tuning was conducted to optimize the random forest model used in this study. The model's performance was evaluated using a 5-fold cross-validation on the training data before being tested on the independent dataset from 2016 to 2022. To achieve the best possible accuracy score for the testing data, a probability threshold tuning was conducted. Finally, the best model was used for the fire susceptibility mapping.

To gain insights into how the model interprets the influence of different features on the predictions, partial dependence plots (PDPs) were calculated for key predictors. This allowed for the visualization of the marginal effect of each feature on the model output while averaging the influence of other features, thus helping to understand their relationship with fire susceptibility.

Additionally, permutation feature importance (PFI) was computed on the training data to assess the contribution of each feature to the model's performance. This technique involves randomly shuffling the values of a feature and measuring the decrease in model accuracy. A greater drop in accuracy indicated higher feature importance for the model - an indicator of which variables may influence wildfire susceptibility most strongly in practice.

3 Results

3.1 Selected Variables

The Relief method selected the following 15 variables: *land cover class*, *proximity to urban areas*, *NDVI*, *distance to rivers*, *surface solar radiation*, *surface sensible heat flux*, *soil water volumetric in the upper most layer*, *LAI low vegetation*, *potential evaporation*, *proximity to agriculture*, *surface pressure*, *eastward wind at 10 meters*, *LAI high vegetation*, *northward wind at 10 meters*, and *proximity to roads*. Of these variables, 7 are related to climate, 4 to vegetation, 3 to human activity, and 1 to topographic conditions. The FFS further narrowed down the variables to *land cover class*, *proximity to urban areas*, *distance to rivers*, *soil water volumetric in the upper most layer*, *LAI low vegetation*, *surface pressure* and *LAI high vegetation*.

The permutation feature importance analysis identified land cover as the most influential variable, causing an accuracy drop of approximately 0.2 when its values were randomly shuffled (figure 3). This was followed by surface pressure and distance to rivers, both leading to an accuracy decrease of around 0.13 (figure 3). Proximity to urban areas, the LAI of low and high vegetation, and soil water resulted in accuracy reductions ranging from 0.08 to 0.1 (figure 3).

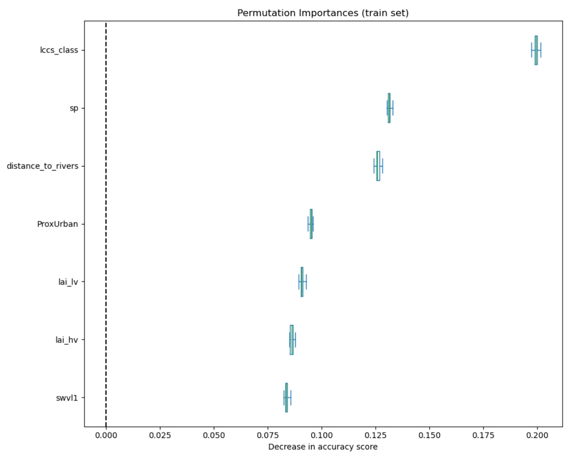


Figure 3: Permutation feature importance calculated on the testing data. Variables include: *lcss_class* (land cover class), *ProxUrban* (proximity to urban areas in degrees of latitude and longitude), *distance_to_rivers* (distance to rivers in degrees of latitude and longitude), *swl1*, *lai_lv* (LAI of low vegetation), *sp* (surface pressure in Pa), and *lai_hv* (LAI of high vegetation).

The partial dependence plot for land cover shows high values around 0.9 for land cover class 11 (rainfed cropland) and class 62 (tree cover, broadleaved, deciduous, closed to open (> 15%)) (figure 4). The land cover classes 120 (broadleaved closed to open shrubland) and 160 (closed to open (100-40%) broadleaved trees on temporarily flooded land) also show values higher than the average (figure 4). The surface pressure exhibits two peaks, at 895 hPa and 980 hPa, with partial dependence reaching up to 0.5 (figure 4). Both distance to rivers and soil water follow a similar pattern, with a slight increase initially up to 0.5, followed by a decrease as the variable values increase (figure 4). Proximity to urban areas and LAI for high vegetation show little to no trend (figure 4), while the LAI for low vegetation demonstrates higher partial dependence between values of 1.4 and 2 (figure 4).

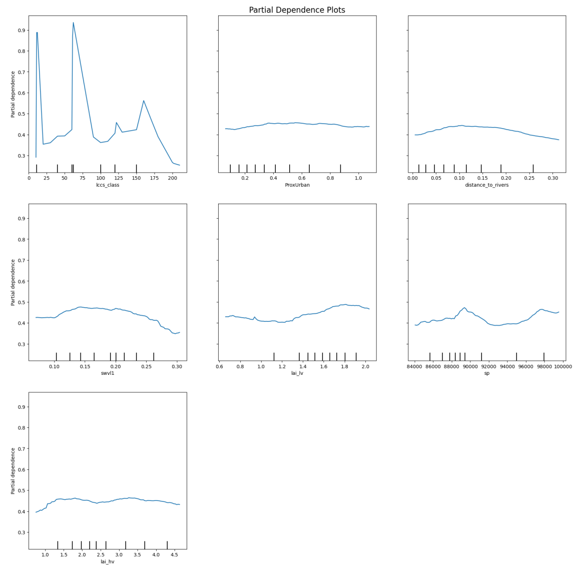


Figure 4: Partial dependence plots calculated on the training data. Variables include: *lcss_class* (land cover class), *ProxUrban* (proximity to urban areas in degrees of latitude and longitude), *distance_to_rivers* (distance to rivers in degrees of latitude and longitude), *swl1*, *lai_lv* (LAI of low vegetation), *sp* (surface pressure in Pa), and *lai_hv* (LAI of high vegetation).

3.2 Hyperparameter and Model evaluation

The grid-search-based hyperparameter tuning yielded the following optimal parameters for the best model: maximum depth = unlimited, minimum samples required at a leaf node = 1, minimum samples required to split an internal node = 3, and number of trees = 1000.

The 5-fold cross-validation resulted in a mean accuracy of 0.873, while the independent test dataset achieved a slightly lower accuracy of 0.846. The probability threshold tuning resulted in an optimal threshold of 0.429 (figure 5), further improving the accuracy on the test data by 0.006, reaching a final accuracy of 0.852.

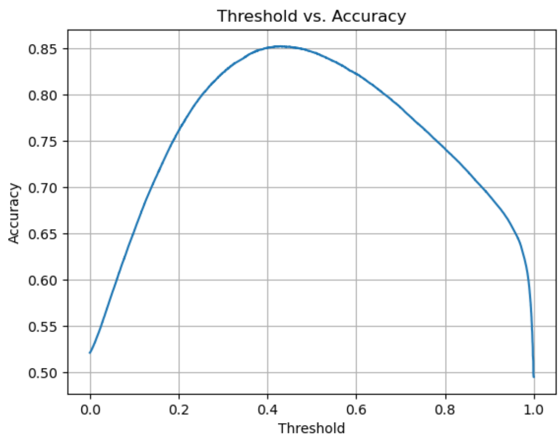


Figure 5: Threshold probability tuning showing the accuracy for each threshold.

The confusion matrix shows that most non-fire pixels were correctly classified. However, 11,500 fire pixels were falsely classified as no fire (figure 6).

3.3 Prediction

The model was applied to predict fire occurrences for both January and July 2022, as respective representatives for wet and dry season. For July, with approximately 1,200 test data points, the model achieved a high accuracy of 0.896. January had only 48 test data points with an accuracy of 0.562. Both prediction examples indicate a high likelihood

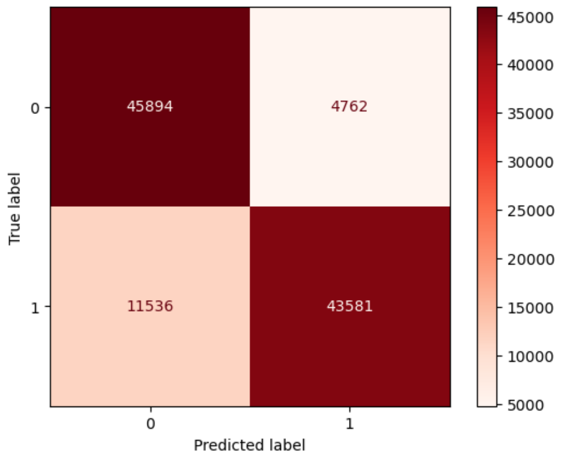


Figure 6: Confusion Matrix of final model prediction after probability threshold tuning.

of fires in the southeastern coastal plains of Tanzania (figure 7). In contrast, the January prediction shows a low fire probability in the plateau grasslands of central Tanzania, while the July prediction reveals a high fire probability in the western Miombo woodlands (figure 7).

4 Discussion

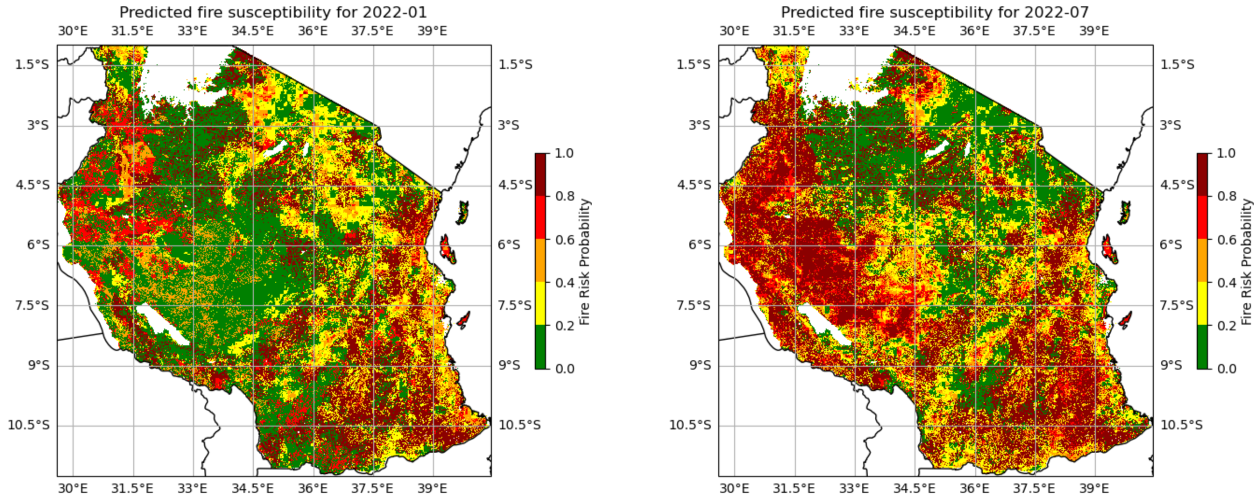
The results of this study showcase the impact by both environmental variables and human activities on wildfires. The model's overall accuracy of 85.2% indicates a strong potential for machine learning techniques to effectively predict wildfire occurrences in this region. However, the varying performance across different seasons and the identified key predictors warrant deeper examination.

4.1 Discussion of Predictor Selection

The method for predictor selection applied in this study proved effective for determining the most important predictors while keeping the model lightweight. Despite computational limitations, both the applied method of predictor selection and the hyperparameter tuning resulted in a model with high accuracy. The selection of land cover class as the most influential predictor aligns with existing literature that emphasizes the importance of vegetation type and land use in wildfire dynamics [20]. Land cover determines the availability of fuels, which represent critical factors in fire ignition and spread [31]. The high partial dependence values for rainfed cropland suggest its particular susceptibility to fires, likely due to agricultural practices such as fire-based post-harvest field clearing [32]. Broad-leaved forests exhibited the highest partial dependence values, likely due to the expansion of fire-prone monocultures like Pinus and Eucalyptus, coupled with - as previously mentioned - close-by agricultural burning [33, 34]. Yet, proximity to agricultural areas was not selected as a predictor, suggesting that other factors may have played a more significant role in fires within broad-leaved forests. Practices such as the negligent use of fire during honey harvesting and other activities may constitute more important factors in these regions [34]. However, there is no sufficient data available to verify these claims. The higher than average partial dependence value of broad leaved shrub-land can be attributed to the natural fire dynamics in this ecosystem.

The permutation feature importance analysis revealed that shuffling the values of land cover led to the largest reduction in model accuracy by a significant margin, indicating that the model heavily relies on this variable as the primary predictor. Land cover encapsulates critical information about vegetation-related variables. Since vegetation acts as primary fuel in wildfires, its composition and structure largely control how a fire behaves and thus determine both the intensity and spread of the flames [20]. Moreover, the type of prevailing vegetation governs moisture retention, and flammability [20]. Therefore, land cover effectively describes these essential aspects of vegetation, explaining its dominance as the most influential predictor in the model.

Another significant factor influencing the model's ability to predict fires was the distance to rivers. The partial dependence increases up to 0.1 degrees of latitude and longitude from



(a) (b)

Figure 7: **(a)** Fire probability map of 01-2022 in the wet season with 48 test points and an accuracy 0.562 **(b)** Fire probability map of 07-2022 in the dry season with 1200 test points and an accuracy of 0.896.

rivers, after which it steadily declines. This indicates that areas closer to rivers are less likely to experience fires than those farther away, as regions near rivers tend to have higher moisture levels, reducing fire risk [25]. The gradual decrease in partial dependence beyond 0.1 degrees in distance to rivers may be attributed to an often simultaneous increase in elevation. In turn, this leads to lower temperatures, further reducing the likelihood of fires. This reflects the combined influence of moisture and elevation on fire risk.

Surface pressure emerged as another significant predictor, with peaks at 895 hPa and 980 hPa correlating with increased fire risk. These peaks likely correspond to specific altitudinal zones in Tanzania due to the inverse relationship between surface pressure and elevation. The peak around 980 hPa represents the coastal plains areas around 300 meters above sea level, which are rich in forests — the land use type experiencing the most fires. The peak at 895 hPa likely reflects higher elevations around 1,000 meters, encompassing highland plateaus that often contain savannas, which represent highly fire-prone regions. This suggests that surface pressure acts as a proxy for altitude-related factors influencing fire occurrence.

Globally calculated fire indices were not selected as predictors, supporting the studys assumption, that they fail to assess actual fire risk at national levels.

4.2 Model Performance Across Seasons

Surface pressure emerged as the only climatic factor selected for fire prediction in the model. However, it primarily operates as a proxy for elevation. Even though dryness is a key factor for fire occurrence, its driving variables, such as temperature or precipitation, remain absent in the model. This absence is likely attributable to the training methodology: non-fire pixels were sampled proportionally to fire pixels for each month and year. Since fires predominantly occur during the dry season and rarely arise in the wet season, the model was mainly trained on data from months showcasing consistently dry conditions. As a result, climatic variables related to dryness did not significantly influence the model's predictions as all training data points correspond to dry conditions.

This is reflected in the model's performance across different seasons. The prediction for July (dry season) yielded an accuracy of 89.6%, whereas the prediction for January (wet season) resulted in a considerably lower accuracy of 56.2%. With an overall accuracy of 85.2%, it transpires that the model is unreliable during the wet season. One might argue that the model's poor performance in months with minimal fire occurrence is inconsequential. However, this lack of responsiveness to short-term climatic variations impedes the model's ability to adapt to changes in weather conditions. Consequently, it may predict high fire susceptibility during dry months even when atypical rainfall has occurred. This represents a significant limitation of the model.

4.3 Implications for Wildfire Management

The findings presented above have significant implications for wildfire management and policy-making in Tanzania. They highlighted vegetation type, coupled with the negligent use of fire by humans, as the strongest factors driving fire risks. This underscores the need for localized fire risk education as a crucial step toward reducing wildfires, particularly for forested areas where they are not a natural occurrence. Educating farmers about the dangers of using fire to clear fields and providing viable alternatives to this method could dramatically lower the frequency of fires in broad-leaved forests. Similarly, promoting safer practices in honey harvesting and charcoal production would further contribute to fire prevention efforts.

Moreover, the study demonstrated the potential of integrating machine learning tools into Tanzania's fire management systems. By offering near-real-time fire risk assessments, authorities can more efficiently allocate resources and inform their population about fire risks in their region, minimizing the impact of wildfires across the country. As criticised at the outset of this study, globally calculated fire risk indices are prone to overlook national-level fire dynamics. The non-selection of these indices by the Relief-F method applied here indicates that more localized fire risk assessments hold greater potential in effective prevention efforts on the national scale.

4.4 Limitations and Future Research

One of the main limitations of this study was the method used to select non-fire pixels. The selection process greatly affects the model's performance and the importance of predictors. In future research, non-fire pixels should be selected across all months, rather than just during dry season. Given the general importance of climate and the particular relevance of dryness in wildfire dynamics, this would allow the model to better capture the influence of climate on fire probability.

A further explanation for the non-selection of climate variables is their coarse spatial and temporal resolution. Such coarseness hampers the model's ability to account for localized environmental factors like moisture gradients, microclimates, and short-term weather events such as heatwaves or rainfall, which are key drivers of fire ignition and spread. This mismatch between fire events and spatial and temporal resolution of climate data led to the loss of important fine-scale dynamics, which reduces the model's predictive accuracy. In future studies, finer-resolution climate data could improve the model's ability to predict fires more accurately by capturing the nuanced interactions between climate, vegetation, and fire risk.

A lack of ground-truth data, particularly regarding human activities known to drive fire occurrence, further restricts the predictive capacities of the model applied. A greater availability of locally recorded data on fire ignition would therefore carry significant potential in improving prediction capabilities.

Additionally, limited computational resources constrained the inclusion of detailed landscape metrics, such as the aggregation and density of land cover types. These metrics, such as forest density, represent important predictors of wildfire occurrence [35]. However, calculating them at a fine spatial resolution using a moving window approach for an area as large as Tanzania over a long time span requires substantial computational power. Future studies could benefit from incorporating these metrics by using high-performance computing resources to handle the demands of such large-scale analyses.

5 Conclusion

This study developed a machine learning model to predict wildfires in Tanzania with an accuracy of 85.2%. Land cover type was identified as the most significant predictor, emphasizing the crucial role of vegetation and human activities in influencing fire risks. The model performed well during dry season but showed reduced accuracy during wet season, primarily due to the high concentration of data points in the dry season. This imbalance likely caused the absence of key dryness-related predictors in the model, limiting its ability to react to climate conditions.

Regardless, integrating such predictive models into fire management systems can enhance early warning capabilities and resource allocation. For example, the findings suggest that educating local communities about fire risks, especially in agricultural and forested areas, is vital for wildfire prevention.

This study demonstrates the potential of machine learning in enhancing wildfire prediction and provides a foundation for better wildfire management strategies in Tanzania. Future research should focus on refining the model by using higher-resolution climate data and including data from all seasons to improve its responsiveness to climatic variations.

Acknowledgments

The author acknowledges the use of ChatGPT for assistance in error handling during coding, grammar and spell checking and finding synonyms during the writing process.

References

[1] M. Sankaran et al. “Determinants of woody cover in African savannas”. In: *Nature* 438.7069 (Dec. 2005), pp. 846–849. DOI: 10.1038/nature04070.

[2] A. C. Staver, S. Archibald, and S. Levin. “Tree cover in sub-Saharan Africa: rainfall and fire constrain forest and savanna as alternative stable states”. In: *Ecology* 92.5 (May 2011), pp. 1063–1072. DOI: 10.1890/10-1684.1.

[3] W. J. Bond, F. I. Woodward, and G. F. Midgley. “The global distribution of ecosystems in a world without fire”. In: *New Phytologist* 165.2 (2005), pp. 525–538. DOI: https://doi.org/10.1111/j.1469-8137.2004.01252.x. eprint: https://nph.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8137.2004.01252.x. URL: https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8137.2004.01252.x.

[4] C. E. Lehmann et al. “Savanna vegetation-fire-climate relationships differ among continents”. In: *Science* 343.6170 (Jan. 2014), pp. 548–552. DOI: 10.1126/science.1247355.

[5] Juli G Pausas and Jon E Keeley. “Wildfires and global change”. In: *Frontiers in Ecology and the Environment* 19.7 (2021), pp. 387–395. DOI: https://doi.org/10.1002/fee.2359. eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1002/fee.2359. URL: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/fee.2359.

[6] Matthew W. Jones et al. "Global and Regional Trends and Drivers of Fire Under Climate Change". In: *Reviews of Geophysics* 60.3 (2022). e2020RG000726 2020RG000726, e2020RG000726. DOI: https://doi.org/10.1029/2020RG000726. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020RG000726. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020RG000726.

[7] Worldometer. *Tanzania Population*. Accessed: 22/09/2024. 2024. URL: https://www.worldometers.info/world-population/tanzania-population/.

[8] Association for Fire Ecology, International Association of Wildland Fire, and The Nature Conservancy. *Reduce Wildfire Risk or we'll continue to pay more for fire disasters*. Accessed: 22/09/2024. 2015. URL: https://www.iawfonline.org/article/reduce-wildfire-risks-or-pay-more-for-fire-disasters/.

[9] J. Makunga and S. Misana. "The Extent and Drivers of Deforestation and Forest Degradation in Masito-Ugalla Ecosystem, Kigoma Region, Tanzania". In: *Open Journal of Forestry* 7 (2017), pp. 285–305. DOI: 10.4236/ojf.2017.72018.

[10] Sally Archibald, A. Carla Staver, and Simon A. Levin. "Evolution of human-driven fire regimes in Africa". In: *Proceedings of the National Academy of Sciences* 109.3 (2012), pp. 847–852. DOI: 10.1073/pnas.1118648109. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1118648109. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1118648109.

[11] Francesco Carta et al. "Advancements in Forest Fire Prevention: A Comprehensive Survey". In: *Sensors* 23.14 (2023). ISSN: 1424-8220. URL: https://www.mdpi.com/1424-8220/23/14/6635.

[12] EUMETSAT. *Fighting Wildfires with Satellite Data in Tanzania*. Accessed: 10/09/2024. n.d. URL: https://www.eumetsat.int/features/fighting-wildfires-satellite-data-tanzania.

[13] Santos Daniel Chicas and Jonas Østergaard Nielsen. "Who are the Actors and What Are the Factors Used in Models to Map Forest Fire Susceptibility? A Systematic Review". In: *Natural Hazards* 114.3 (2022), pp. 2417–2434.

[14] Ayn Shlisky et al. "Overview: Global Fire Regime Conditions, Threats, and Opportunities for Fire Management in the Tropics". In: *Tropical Fire Ecology: Climate Change, Land Use, and Ecosystem Dynamics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 65–83. ISBN: 978-3-540-77381-8. DOI: 10.1007/978-3-540-77381-8_3. URL: https://doi.org/10.1007/978-3-540-77381-8_3.

[15] J. Muñoz Sabater. *ERA5-Land monthly averaged data from 1950 to present*. 2019. DOI: 10.24381/cds.68d2bb30. URL: https://doi.org/10.24381/cds.68d2bb30.

[16] NASA FIRMS. *MODIS Collection 6 Hotspot / Active Fire Detections MCD14ML*. Distributed by NASA FIRMS. DOI: 10.5067/FIRMS/MODIS/MCD14ML. URL: https://earthdata.nasa.gov/firms.

[17] Copernicus Climate Change Service. *Land cover classification gridded maps from 1992 to present derived from satellite observation*. 2019. DOI: 10.24381/cds.006f2c9a. URL: https://doi.org/10.24381/cds.006f2c9a.

[18] Binh Thai Pham et al. "Performance Evaluation of Machine Learning Methods for Forest Fire Modeling and Prediction". In: *Symmetry* 12.6 (2020). ISSN: 2073-8994. DOI: 10.3390/sym12061022. URL: https://www.mdpi.com/2073-8994/12/6/1022.

[19] Sanjeev Sharma and Puskar Khanal. "Forest Fire Prediction: A Spatial Machine Learning and Neural Network Approach". In: *Fire* 7.6 (2024). ISSN: 2571-6255. DOI: 10.3390/fire7060205. URL: https://www.mdpi.com/2571-6255/7/6/205.

[20] E. L. Loudermilk, J. J. O'Brien, S. L. Goodrick, et al. "Vegetation's influence on fire behavior goes beyond just being fuel". In: *Fire Ecology* 18 (2022), p. 9. DOI: 10.1186/s42408-022-00132-9. URL: https://doi.org/10.1186/s42408-022-00132-9.

[21] NASA Jet Propulsion Laboratory. *Shuttle Radar Topography Mission (SRTM) 30 Meter Digital Elevation Model*. 2000. DOI: 10.5067/MEaSUREs/SRTM/SRTMGL1.003. URL: https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003.

[22] S. Sharma et al. "Characterizing Topographic Influences of Bushfire Severity Using Machine Learning Models: A Case Study in a Hilly Terrain of Victoria, Australia". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16 (2023), pp. 2791–2807. DOI: 10.1109/JSTARS.2023.3249643.

[23] P. Nyman et al. "Quantifying the effects of topographic aspect on water content and temperature in fine surface fuel". In: *International Journal of Wildland Fire* 24 (2015), pp. 1129–1142. DOI: 10.1071/WF14195.

[24] OpenStreetMap contributors. *Street and railway data*. Data retrieved from OpenStreetMap and licensed under the Open Database License (ODbL). 2024. URL: https://www.openstreetmap.org.

[25] A.T. Tesfaye and D.T. Defersha. "Contribution to the Land Suitability Analysis for Potential Surface Irrigation Development Using Remote Sensing and GIS-MCE of the Soroka Watershed, Northwestern Ethiopia". In: *Sustainable Water Resources Management* 10 (2024), p. 51. DOI: 10.1007/s40899-024-01033-2.

[26] K. Didan. *MOD13A3.061 Terra Vegetation Indices Monthly Global 1km*. 2015. DOI: 10.5067/MODIS/MOD13A3.061. URL: https://doi.org/10.5067/MODIS/MOD13A3.061.

[27] Copernicus Climate Change Service. *Fire danger indices historical data from the Copernicus Emergency Management Service*. Accessed on DD-MMM-YYYY. 2019. DOI: 10.24381/cds.0e89c522. URL: https://doi.org/10.24381/cds.0e89c522.

[28] Kenji Kira and Larry A. Rendell. "A Practical Approach to Feature Selection". In: *ML Workshop*. 1992. URL: https://api.semanticscholar.org/CorpusID:33296374.

[29] Igor Kononenko. "Estimating Attributes: Analysis and Extensions of RELIEF". In: *European Conference on Machine Learning*. 1994. URL: https://api.semanticscholar.org/CorpusID:8190856.

[30] Ryan J. Urbanowicz et al. "Relief-Based Feature Selection: Introduction and Review". In: *Journal of Biomedical Informatics* 85 (2018), pp. 189–203. ISSN: 1532-0464. DOI: https://doi.org/10.1016/j.jbi.2018.07.014. URL: https://www.sciencedirect.com/science/article/pii/S1532046418301400.

[31] Susan J. Prichard et al. "Fuels and Consumption". In: *Wildland Fire Smoke in the United States: A Scientific Assessment*. Ed. by David L. Peterson, Sarah M. McCaffrey, and Toral Patel-Weynand. Cham: Springer International Publishing, 2022, pp. 11–49. ISBN: 978-3-030-87045-4. DOI: 10.1007/978-3-030-87045-4_2. URL: https://doi.org/10.1007/978-3-030-87045-4_2.

[32] JOSEPH M DiTomaso and Douglas W Johnson. "The use of fire as a tool for controlling invasive plants". In: *Cal-IPC publication* 1 (2006), p. 56.

[33] Antti Erkkilä et al. "Ambiguity and forest-based bioeconomy: The case of forest fires in the Southern Highlands of Tanzania". In: *Forest Policy and Economics* 160 (2024), p. 103159. ISSN: 1389-9341. DOI: https://doi.org/10.1016/j.forpol.2024.103159. URL: https://www.sciencedirect.com/science/article/pii/S1389934124000121.

[34] Samora Macrice Andrew. "Drivers, trends and management of forest plantation fires in Tanzania". In: *Trees, Forests and People* 10 (2022), p. 100355. ISSN: 2666-7193. DOI: https://doi.org/10.1016/j.tfp.2022.100355. URL: https://www.sciencedirect.com/science/article/pii/S2666719322001613.

[35] Cristobal Pais et al. "Deep fire topology: Understanding the role of landscape spatial patterns in wildfire occurrence using artificial intelligence". In: *Environmental Modelling Software* 143 (2021), p. 105122. ISSN: 1364-8152. DOI: https://doi.org/10.1016/j.envsoft.2021.105122. URL: https://www.sciencedirect.com/science/article/pii/S1364815221001651.

Appendix

Source	**Predictor**
Copernicus Climate Change Service	Land cover classification system (lccs_class)
ERA5	10-meter U-component of wind (u10)
	10-meter V-component of wind (v10)
	2-meter dewpoint temperature (d2m)
	2-meter temperature (t2m)
	Potential evaporation (pev)
	Skin temperature (skt)
	Soil temperature level 1 (stl1)
	Surface latent heat flux (slhf)
	Surface solar radiation (ssr)
	Surface pressure (sp)
	Surface sensible heat flux (sshf)
	Evaporation (e)
	Total precipitation (tp)
	Volumetric soil water layer 1 (swvl1)
	Relative humidity (rel_hum)
	Vapor pressure (vapPres)
	Vapor pressure deficit (VPD)
	High vegetation leaf area index (lai_hv)
	Low vegetation leaf area index (lai_lv)
	Surface roughness coefficient (src)
MODIS	Normalized Difference Vegetation Index (NDVI)
SRTM	Digital Elevation Model (DEM)
	Aspect
	Topographic Wetness Index (TWI)
	Hillshade index (hli)
	Distance to rivers
	Slope
OpenStreetMap	Proximity to urban areas (ProxUrban)
	Proximity to agricultural areas (ProxAgri)
	Proximity to roads (ProxiRoads)
Copernicus Emergency Management Service	Build-up index (fbupinx)
	Fire danger index (fdimrk)

Table 1: Data Sources. Copernicus Climate change Service [27], ERA5 [15], MODIS [26], SRTM [21], OpenStreetMap [24], Copernicus Emergency Management Service [27]