Received: 28 December 2025 | Revised: 9 June 2026 | Accepted: 17 June 2026 | Published online: 28 July 2026

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2026, Vol. 00(00) 1–19
DOI: 10.47852/bonviewJDSIS62028499

BON VIEW PUBLISHING

Spatiotemporal Analysis of Climate–Forest Fire Correlation in Peninsular Malaysia Using SARIMAX and Ensemble Models

Grace Kian Hwai Ling[1] and Weng Howe Chan[1,*]

[1] *Faculty of Computing, University Technology Malaysia, Malaysia*

Abstract: Forest fires in Peninsular Malaysia continue to threaten biodiversity, public health, and economic activity, with risks most evident in peat swamp forests where prolonged dry periods increase fuel flammability. This study analyzes 22 years of fire occurrence (2001–2023) across peat swamp forest regions in Pahang, Selangor, Johor, and Terengganu to examine long-term climate–fire relationships and to develop predictive tools for fire-risk assessment. The modeling framework was organized into three sequential stages. Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (SARIMAX) was applied for temporal modeling, while an ensemble framework combining Generalized Linear Models and Maximum Entropy estimated fire occurrence probability. A hybrid system subsequently linked SARIMAX-derived climatic projections with ensemble classifiers to generate forward-looking estimates. The results reveal a recurring seasonal pattern, characterized by a clear March peak and noticeably higher fire activity during drought years, including 2005 and 2014. SARIMAX produced the strongest temporal accuracy (average Mean Absolute Scaled Error = 0.1902), demonstrating substantially lower forecast error (around 81%) than a simple seasonal repetition benchmark. Spatial probability modeling showed strong climatic separability at the district–monthly scale (area under the curve = 1; Continuous Boyce Index = 0.67). Projection outputs suggest a gradual redistribution of hotspot intensity from Sabak Bernam toward inland districts such as Pekan, although these estimates remain conditional on projected climatic patterns. Overall, the integrated framework offers a climate-informed foundation for operational monitoring and longer-term management of peat swamp forest fires in Peninsular Malaysia.

Keywords: climate–fire dynamics, forest fire forecasting, SARIMAX, ensemble modeling, spatiotemporal analysis

1. Introduction

Forest fires in Peninsular Malaysia present substantial risks to ecological stability, public health, and economic activities [1, 2]. These threats are particularly evident in peat swamp forests, where deep organic soil layers become highly combustible when groundwater levels decline during extended dry periods [3]. Peatlands function as major carbon reservoirs, and their degradation contributes to increased carbon emissions while weakening ecosystem resilience [4]. Severe haze episodes, including the 2019 event, demonstrate the wide-ranging impacts of peat-related fires on air quality, respiratory health, transportation systems, and agricultural productivity [5].

Across the region, fluctuations in climatic conditions play a central role in shaping forest fire behavior. El Niño events intensify drought conditions by suppressing rainfall and elevating temperature levels [3]. The 1997–1998 El Niño contributed to extensive burning in Sabah, illustrating the severe consequences of prolonged hydroclimatic stress [4]. Within tropical peat-dominated systems, rising temperature accelerates evapotranspiration, whereas reduced precipitation lowers both surface and subsurface moisture availability. Hydrological indicators reflect deeper peat moisture dynamics, while vegetation metrics signal canopy stress prior to fire occurrence [6]. Together, these climatic processes regulate ignition likelihood and fire intensity.

Evidence from both global and regional assessments points to increasingly fire-conducive climatic conditions across Southeast Asia [7, 8], accompanied by a warming trend of approximately 0.25 °C per decade in Malaysia [9]. Declining moisture availability at both surface and subsurface levels increases fuel flammability. These interacting processes highlight the importance of understanding long-term climate–fire relationships, particularly within peat swamp ecosystems, where hydrological balance directly influences fire susceptibility.

Over time, forest fire research has moved beyond descriptive analyses toward more advanced modeling approaches. Time-series models such as Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) are widely applied to examine fire trends and seasonal recurrence patterns [10, 11]. Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (SARIMAX) extends this structure by incorporating exogenous climatic variables, allowing environmental drivers to directly influence temporal fire dynamics. Machine-learning approaches, particularly ensemble modeling frameworks that integrate predictions

***Corresponding author:** Weng Howe Chan, Faculty of Computing, University Technology Malaysia, Malaysia. Email: cwenghowe@utm.my

from multiple algorithms (e.g., Generalized Linear Models [GLM] and Maximum Entropy [MaxEnt]) [12, 13], estimate fire occurrence probability by learning relationships between observed fire occurrences and climatic or environmental predictors. Although these approaches have been applied in other tropical regions, their adoption in Peninsular Malaysia remains comparatively limited.

This study focuses specifically on climatic and environmental drivers of forest fire occurrence using a long-term remote-sensing dataset comprising 19 climate-related variables. Although peatland fires in Southeast Asia may also be influenced by anthropogenic factors such as agricultural burning, land-use change, and slash-and-burn activities, these variables were not included in the available dataset used in this study. Consequently, the modeling framework was designed to evaluate climate–fire relationships and develop climate-driven forecasting models. The integration of anthropogenic and land-use variables remains an important direction for future research.

Existing Malaysian studies frequently rely on datasets spanning fewer than 10 years and often emphasize rainfall and temperature variables [14]. Hydrological indicators and vegetation stress metrics that are critical within peat swamp ecosystems are less consistently incorporated. Furthermore, temporal forecasting and spatial fire-risk modeling are commonly conducted separately, limiting the translation of climatic projections into spatially explicit fire-risk assessments.

To overcome these limitations, the present study examines 22 years of forest fire occurrences together with 19 climatic variables across peat swamp forests in Pahang, Selangor, Johor, and Terengganu. The extended dataset captures multiple El Niño-Southern Oscillation (ENSO) phases, enabling clearer differentiation between recurring seasonal behavior and episodic climatic anomalies. The research integrates SARIMAX for temporal forecasting with ensemble modeling combining GLM and MaxEnt for fire occurrence probability estimation. A hybrid forecasting framework links SARIMAX-derived climatic projections with ensemble classifiers to estimate future fire probabilities and projected monthly fire counts.

The objectives of this study are to (i) provide a long-term assessment of climate–fire dynamics in peat swamp forests of Peninsular Malaysia, (ii) evaluate predictive performance of integrated temporal and probabilistic models, and (iii) support fire-risk management through a decision-support dashboard. By combining extended climatic analysis with integrated modeling, the study contributes a structured climate-driven fire forecasting framework tailored to peat swamp environments in Peninsular Malaysia.

2. Theoretical Framework

The study's theoretical foundation is grounded in established climate–fire relationships and predictive modeling frameworks that capture both temporal dynamics and climate-driven fire susceptibility.

2.1. Climatic drivers of forest fire behavior

Forest fire behavior is regulated by complex atmospheric and hydrological processes that control fuel moisture and ignition potential [1, 15, 16]. Rising temperature accelerates evapotranspiration, which dries out surface vegetation and deep organic peat layers, whereas precipitation serves to replenish these vital moisture reserves [3, 14]. Extended dry conditions are captured through drought indices, while hydrological measures reflect subsurface moisture dynamics within peat soils [17, 18]. Remote-sensing vegetation indices indicate canopy stress and declining greenness, conditions that often precede fire events [6, 17].

2.2. Temporal modeling framework

Time-series models provide a structured approach for analyzing monthly fire occurrence while accounting for seasonal cycles and long-term trends. ARIMA decomposes time-series observations into autoregressive (AR), differencing, and moving-average (MA) components. SARIMA extends this structure by incorporating seasonal parameters to capture recurring interannual patterns [11]. SARIMAX further integrates exogenous climatic predictors, enabling environmental variables to directly influence temporal fire dynamics. This framework is particularly suitable for modeling recurring seasonal peaks and climate-driven variability observed in Peninsular Malaysia.

2.3. Spatial predictive modeling

Spatial predictive modeling estimates the probability of fire occurrence under specific climatic conditions. GLM applies a binomial logistic regression framework to quantify relationships between predictors and fire presence, offering a robust approach for non-normally distributed response variables [19]. The MaxEnt estimates climatic suitability using presence–background comparison and captures nonlinear response behavior [20]. These models are frequently integrated into ensemble frameworks using area under the curve (AUC)-weighted averaging to improve predictive accuracy [12, 13]. In the present study, AUC-based weighting was adopted to assign greater influence to models demonstrating stronger discrimination capability between fire and non-fire conditions. This approach provides a straightforward and widely applied mechanism for combining predictive outputs while maintaining consistency with the evaluation framework used throughout the study. Such approaches generate probabilistic fire-risk surfaces reflecting climatic separability between fire and non-fire conditions at the district–monthly scale. Visualization through Power BI further supports the interpretation and operational application of these modeling outputs for fire managers.

3. Methodology

An integrated methodological framework was adopted to investigate long-term climate–fire relationships across 36 districts in Peninsular Malaysia. All analyses were conducted using the forest fire dataset developed by Chew et al. [17], which consolidates 22 years of fire incident records with 19 climatic and environmental variables derived from remote sensing and TerraClimate repositories.

The methodological workflow comprises data preparation, exploratory analysis, and multicollinearity screening using variance inflation factor (VIF) and tolerance (TOL) diagnostics to identify and remove highly correlated predictors, thereby ensuring model stability and reducing coefficient inflation [12]. Given the strong interrelationships commonly observed among tropical climatic variables, established thresholds of $VIF \leq 10$ and $TOL \geq 0.1$ were adopted to minimize predictor redundancy while retaining meaningful climatic information. The refined predictor set was subsequently used for temporal modeling using SARIMAX, probabilistic fire occurrence modeling using a weighted ensemble of GLM and MaxEnt, hybrid forecasting for 10-year projections, and interactive dashboard visualization for decision support. Each stage was designed to maintain statistical robustness while supporting interpretability and practical relevance for regional fire-risk mitigation.

3.1. Research design

The research design integrates descriptive analysis, temporal forecasting, and probabilistic modeling to investigate forest fire behavior at the district scale. The framework combines:

1) Seasonal time-series modeling (SARIMAX) to analyze monthly fire counts and climatic drivers;
2) Classification models (GLM and MaxEnt) to estimate climate-based fire occurrence probability;
3) A hybrid SARIMAX–ensemble framework to project future fire probability and monthly fire counts.

Through this combination, the framework captures temporal forecasting of fire frequency alongside spatial estimation of climatic suitability for fire occurrence.

To ensure reliable model evaluation, a forward-chaining temporal train–test split was applied. Monthly observations from January 2001 to December 2018 (approximately 80%) were used for training, while observations from January 2019 to December 2023 (approximately 20%) were reserved for testing. This preserves chronological order and reflects realistic forecasting conditions, where future fire activity is predicted using only historical information.

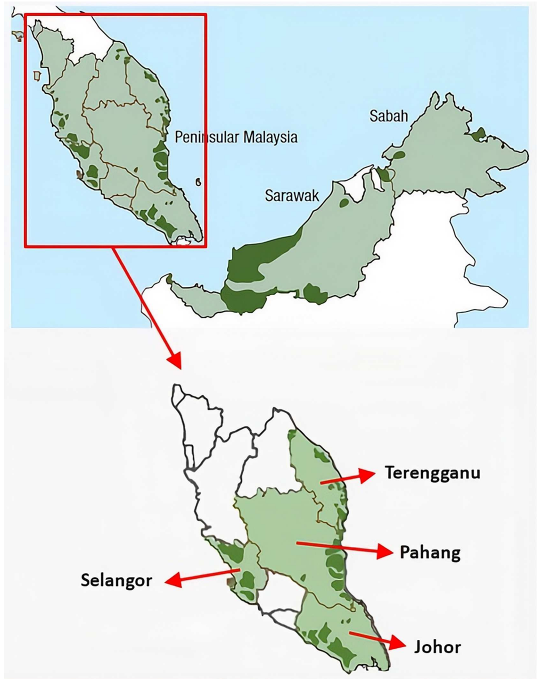
For probabilistic models, climatic predictors were temporally aligned with the response variable to avoid contemporaneous bias. Model performance therefore reflects genuine out-of-sample predictive ability.

Interpretability considerations varied across modeling frameworks. For temporal models (ARIMA, SARIMA, and SARIMAX), diagnostic procedures including Augmented Dickey–Fuller stationarity testing, differencing assessment, and information-criterion–based parameter optimization provided insight into temporal dependence structures during model development. For GLM, MaxEnt, and ensemble models, evaluation focused on predictive performance using receiver operating characteristic–area under curve (ROC–AUC), True Skill Statistic (TSS), Continuous Boyce Index (CBI), and accuracy metrics to assess discrimination ability and reliability. This design emphasized comparative predictive assessment while maintaining methodological consistency across statistical and machine-learning approaches.

3.2. Study area

The study focuses on peat swamp forest regions in four states of Peninsular Malaysia: Pahang, Selangor, Johor, and Terengganu (Figure 1) [21]. These states contain extensive peatland areas characterized by deep organic soils, high carbon accumulation, and seasonal water-table variation.

Figure 1
Peat swamp forest in Peninsular Malaysia



Peatlands are highly sensitive to prolonged drying. When groundwater levels decline, peat layers become exposed and highly flammable, particularly during periods of atmospheric moisture stress [3]. Seasonal drought therefore plays a critical role in increasing fire susceptibility.

Thirty-six administrative districts were used as spatial units. District boundaries were derived from official administrative shapefiles, ensuring consistency between climatic aggregation and fire incident records.

3.3. Data sources

All analyses were based on the forest fire dataset (Figure 2) [17]. Fire records were obtained from the MODIS MCD64A1 Burned Area product covering the period 2001–2023. These records provide spatial coordinates, dates, and fire incident counts.

Climatic and environmental variables were derived from high-resolution remote-sensing products (e.g., MOD11A2.061 for land surface temperature [LST] and MOD13Q1.061 for vegetation) and the TerraClimate repository (Figure 3). Nineteen indicators were included, covering:

1) Atmospheric/Meteorological Variables: minimum temperature (tmnn), maximum temperature (tmnx), LST, precipitation accumulation (pr), wind speed (vs), downward surface shortwave radiation (srad), vapor pressure (vap), vapor pressure deficit (vpd);
2) Drought Indicators: Palmer Drought Severity Index (PDSI), Keetch–Byram Drought Index (KBDI);
3) Hydrological Variables: soil moisture (soil), runoff (ro), climate water deficit (def), snow water equivalent (SWE);
4) Evapotranspiration Variables: actual evapotranspiration (aet), reference evapotranspiration (pet);
5) Vegetation and Land Surface Characteristics: Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), land cover (LC).

All variables were spatially aggregated to the district level and structured as monthly observations aligned with fire incident data.

For probabilistic modeling, fire occurrence was defined as a binary variable:

Figure 2
Forest fire dataset

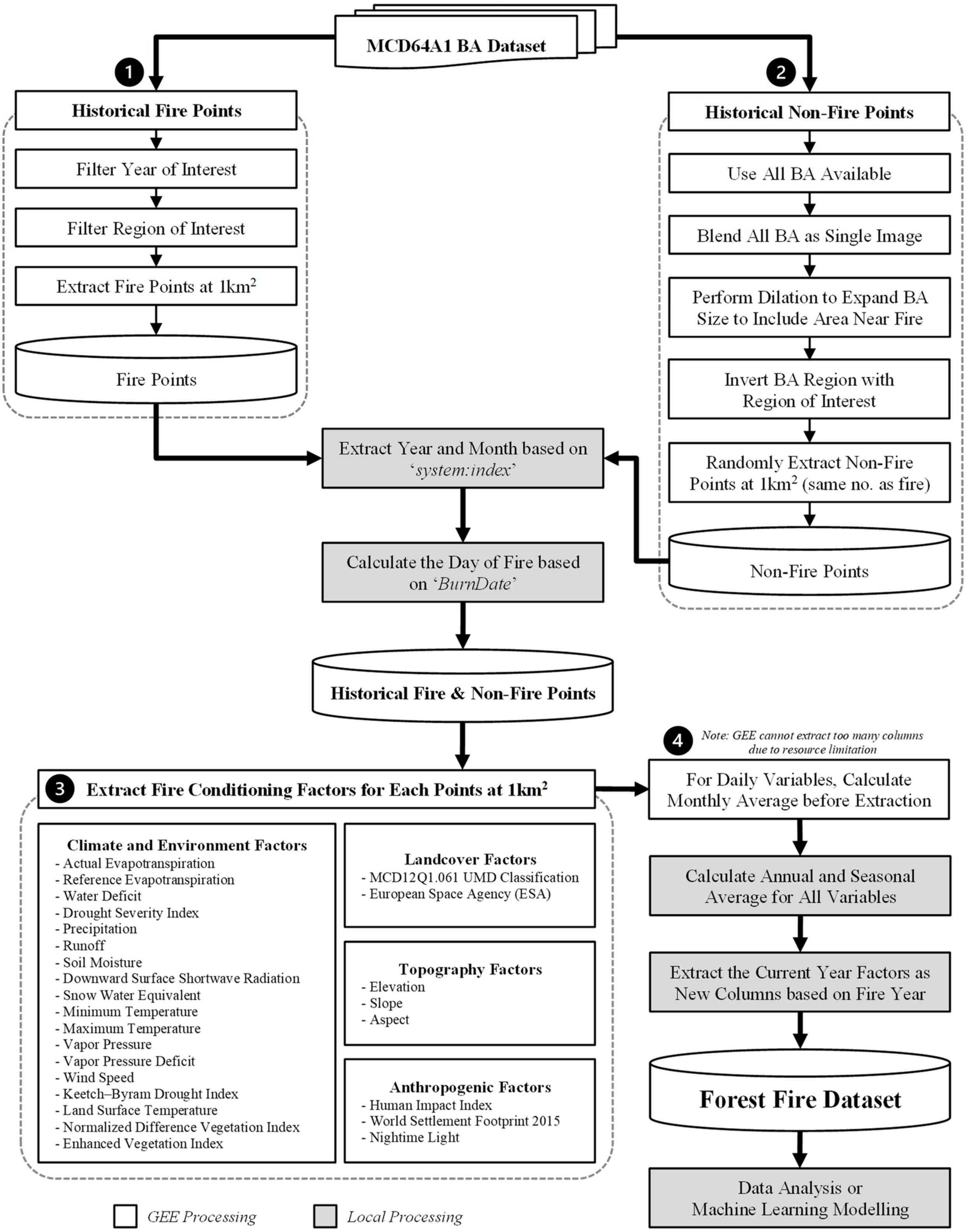


Figure 3
Dataset features and descriptions

Feature Name	Description	Feature Name	Description
system:index	System-generated from MCD64A1	current0101_hii_annual	Human Impact Index
longitude	Longitude Coordinate of Fire Points	current0101_average_annual_nighttime	Nighttime Brightness
latitude	Latitude Coordinate of Fire Points	current_aet_annual	Actual Evapotranspiration
fire	Fire Occurrence (binary class)	current_def_annual	Climate Water Deficit
date	Date from Administrative Boundaries refer to the Shape	current_pdsi_annual	Palmer Drought Severity Index
ADM0_PCODE	Country code	current_pet_annual	Reference Evapotranspiration
ADM1_PCODE	Administrative level 1 code	current_pr_annual	Precipitation Accumulation
ADM2_PCODE	Administrative level 2 code	current_ro_annual	Runoff
Shape_Leng	Shape Length (from MCD64A1)	current_soil_annual	Soil Moisture
elevation	Elevation	current_srad_annual	Downward Surface Shortwave Radiation
slope	Slope (derived from DEM)	current_swe_annual	Snow Water Equivalent
aspect	Aspect (derived from DEM)	current_tmnn_annual	Minimum Temperature
hillshade	Hill Shade	current_tmnx_annual	Maximum Temperature
ADM0_EN	Country Name	current_vap_annual	Vapor Pressure
ADM1_EN	Administrative level 1 name	current_vpd_annual	Vapor Pressure Deficit
ADM2_EN	Administrative level 2 name	current_vs_annual	Wind Speed at 10 m
validOn	Validation Date from Administrative Boundaries refer to the Shape	current_EVI_annual	Enhanced Vegetation Index
Shape_Area	Shape Area (from MCD64A1)	current_NDVI_annual	Normalized Difference Vegetation Index
BurnDate	Date in 0–365 (from MCD64A1)	current_LST_annual	Land Surface Temperature
year	Year of Fire Observation	current_KBDI_annual	Keetch-Byram Drought Index
month	Month of Fire Observation	current0101_LC_Type2_annual	Land Cover Classification of UMD (Numeric)
day	Day of Fire Observation	current0101_LC_Type2_annual_classname	Land Cover Classification of UMD (Class Name)

1) 1 if at least one fire incident occurred within a district during a given month;
2) 0 otherwise.

This definition reflects the operational objective of distinguishing fire-prone versus non-fire climatic conditions at the district–monthly scale.

3.4. Exploratory data analysis

Prior to model development, exploratory analysis was conducted to examine temporal trends and the associations between climatic variables and fire occurrences. A set of 19 annual climatic variables was compared between fire and non-fire periods using boxplots (Figure 4). This visualization allowed examination of medians, interquartile ranges, and overall variability, highlighting both central tendency and spread of the data for each variable.

Comparative analysis indicated that fire months generally exhibited higher median values of drought- and temperature-related variables, including KBDI, PET, and maximum temperature. In contrast, variables representing moisture and vegetation conditions, such as precipitation, soil moisture, and NDVI, tended to have lower median values during fire periods. Wider interquartile ranges were observed for certain variables (e.g., KBDI, climatic water deficit, and PDSI) during fire months, reflecting greater variability and the influence of extreme dry conditions on fire susceptibility. Additional atmospheric variables, including wind speed and vapor pressure, exhibited higher median values during fire periods, further indicating the role of moisture–energy interactions in influencing ignition conditions.

Figure 4
Outliers' boxplot

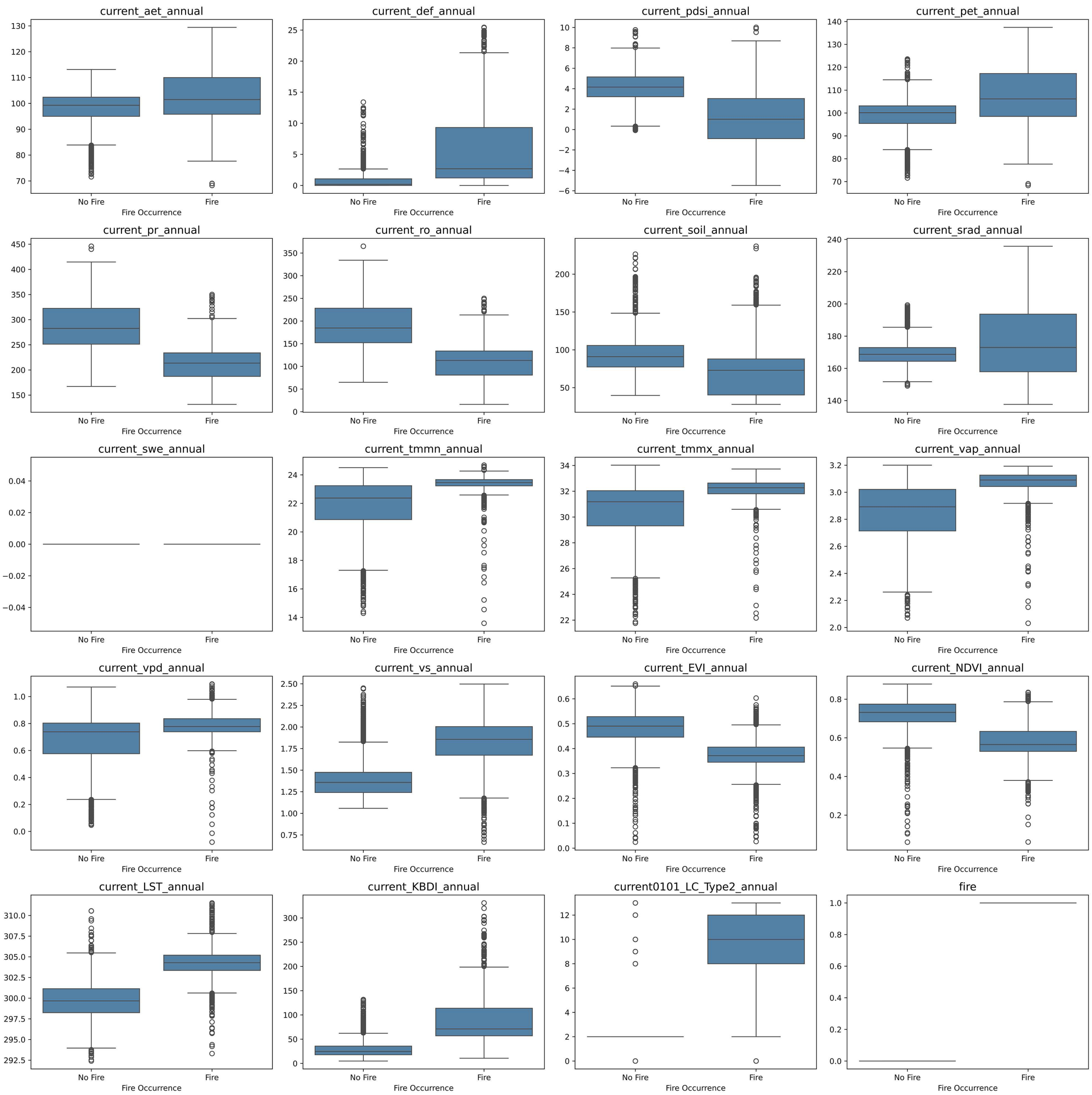
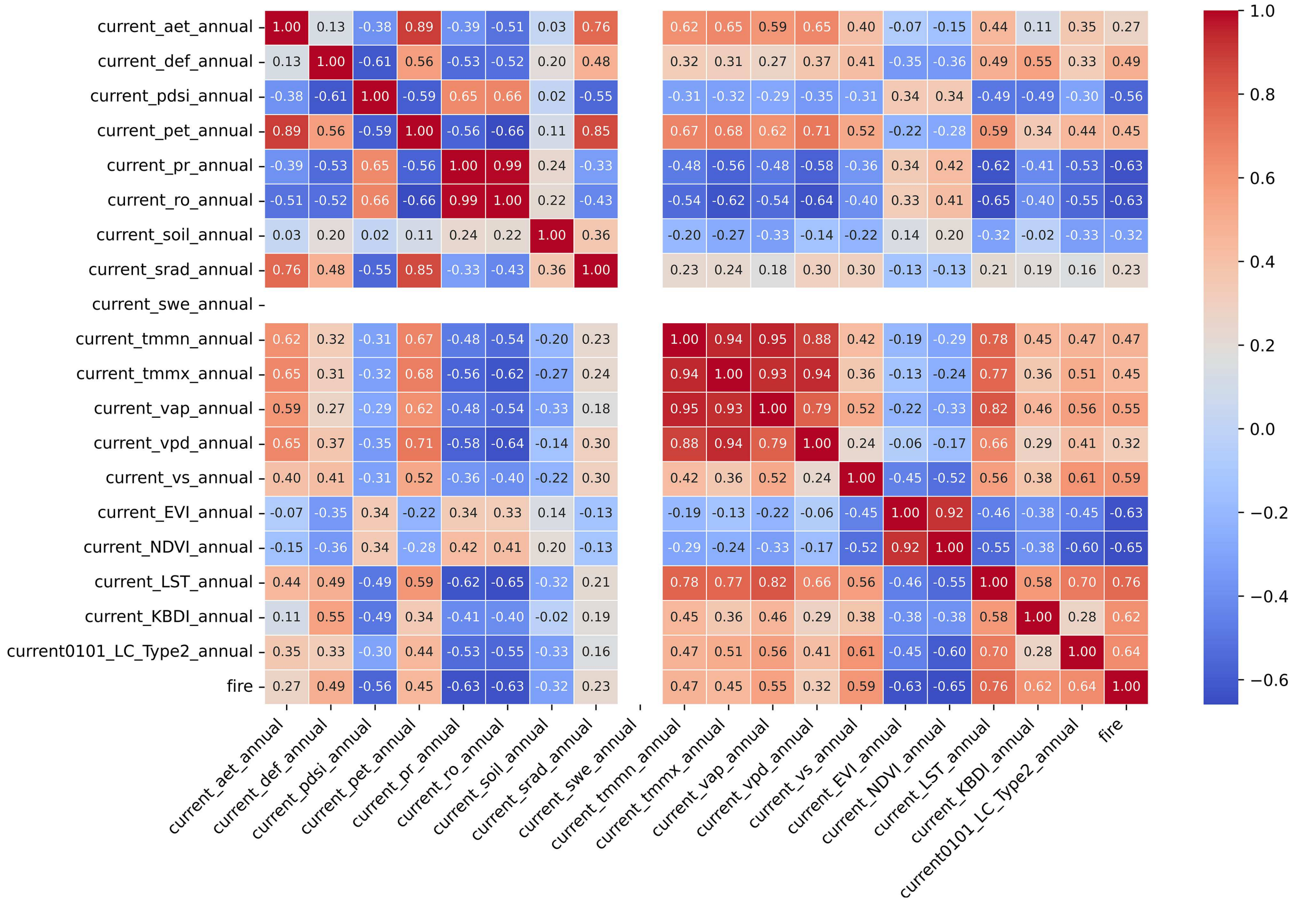


Figure 5
Climatic variables' correlations



Correlation analysis (Figure 5) revealed strong climatic associations with fire occurrence. LST exhibited the strongest positive relationship ($r = 0.76$), indicating the influence of surface heat accumulation on ignition potential. Drought-related indices, particularly the KBDI ($r = 0.62$), also showed strong positive correlations, reflecting the role of prolonged moisture deficit in increasing fuel flammability.

Conversely, vegetation and hydrological indicators demonstrated significant negative relationships. Negative NDVI ($r = -0.65$) and runoff ($r = -0.63$) suggest that reduced vegetation health and lower surface moisture are strongly linked to fire incidence. The PDSI ($r = -0.56$) further reinforces the contribution of atmospheric dryness to fire susceptibility.

Additional atmospheric variables, including vapor pressure, vapor pressure deficit, and wind speed, displayed moderate associations with fire activity, highlighting the combined influence of moisture stress and aerodynamic conditions on fire ignition and spread.

Time-series visualization from 2001 to 2023 (Figure 6) highlighted pronounced seasonal and interannual fire patterns. Fire activity showed consistent concentration during the late dry season, with peaks typically occurring in March. Elevated fire incidence was particularly evident during major El Niño periods, including 2005 and 2014, reflecting the influence of large-scale climatic anomalies on regional fire regimes.

Normalized trend comparisons further revealed strong temporal alignment between fire peaks and increases in heat- and drought-related variables, including LST and KBDI. In contrast, hydrological and vegetation indicators such as precipitation, runoff, NDVI, and EVI exhibited inverse patterns, declining during major fire years. These synchronous fluctuations illustrate the compound influence of heat accumulation, atmospheric dryness, and vegetation stress on fire occurrence.

Years with relatively low fire activity corresponded to wetter climatic conditions, characterized by higher precipitation, stronger vegetation greenness, and reduced drought severity. Collectively, these seasonal and interannual dynamics indicate strong temporal structure in fire occurrence, supporting the application of seasonal time-series modeling frameworks in subsequent analyses.

3.5. Data preprocessing

Data preprocessing (Figure 7) was carried out to improve consistency, completeness, and overall suitability for modeling.

3.5.1. Cleaning and filtering

The original dataset comprised 11,279 fire occurrence records alongside more than 7000 feature columns, including extensive metadata, geographic references, and administrative identifiers. Prior to missing-value treatment, irrelevant attributes were removed to streamline the analytical structure. These included geographic coordinates (e.g., *longitude, latitude*), administrative codes (e.g., *ADM1_PCODE, ADM2_PCODE, ADM2_REF*), shape metadata (e.g., *Shape_Area, Shape_Leng*), temporal system fields (e.g., *BurnDate, validOn*), and artifact index columns (e.g., *Unnamed: 0*). Eliminating these non-analytical fields improved

Figure 6
Normalized variables and forest fire occurrence

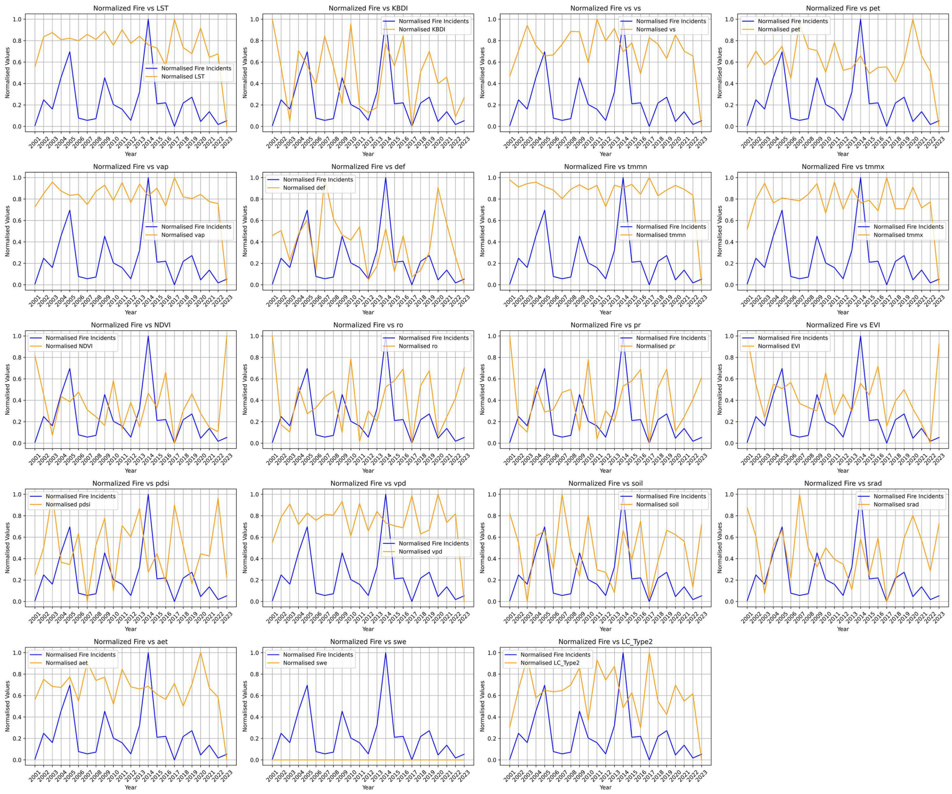
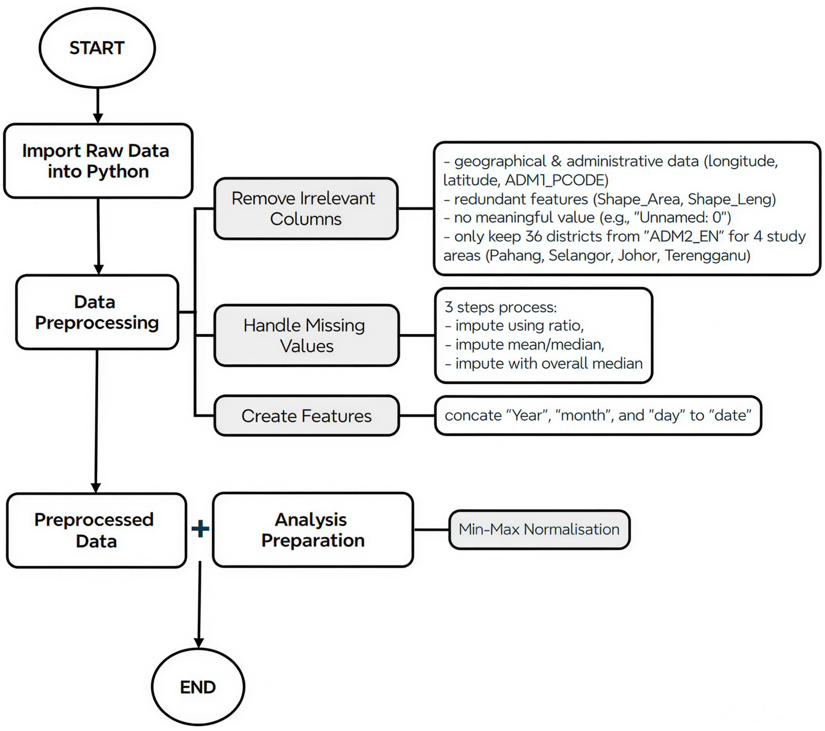


Figure 7
Data preparation flowchart



dataset manageability and ensured that subsequent modeling focused on variables directly associated with forest fire dynamics.

District filtering was subsequently applied using the *ADM2_EN* field to retain only the 36 districts located within the selected study states of Pahang, Selangor, Johor, and Terengganu.

3.5.2. Missing value handling

Missing values were addressed using a structured hierarchical approach:

1) Categorical standardization: Missing entries in *ADM2_EN* were replaced with "Unknown" to preserve grouping.
2) Ratio-based imputation: Related numeric variables were imputed using median-derived ratios from closely associated features.
3) District-level imputation: Remaining missing values were filled using mean or median statistics within each *ADM2_EN* group, where the median was applied when the mean–median difference was ≥ 1.
4) Global fallback imputation: Any residual gaps were replaced using overall column medians.

The structured imputation strategy was intended to preserve ecological relationships while improving data completeness.

3.5.3. Normalization and time structuring

Climatic variables were normalized using min–max scaling to address differences in measurement scales. Variables with large numeric ranges, such as precipitation, were rescaled alongside bounded indicators such as vegetation indices to ensure comparability during modeling. The *MinMaxScaler* algorithm was applied to standardize all numerical features within a 0–1 range, reducing potential bias arising from unequal feature magnitudes and improving model interpretability.

Temporal fields (*year, month, day*) were first validated and converted into numeric formats, with non-numeric entries coerced into missing values to ensure data consistency. These components were subsequently concatenated into a unified datetime variable and aggregated into a monthly time index spanning January 2001 to December 2023. The dataset was then reorganized into a continuous monthly sequence for each district, fulfilling the temporal structuring requirements of SARIMAX modeling and supporting seasonal and trend analysis.

After preprocessing, cleaning, and filtering procedures, the final analytical dataset comprised 7272 complete district–monthly records.

3.6. Temporal modeling using SARIMAX

Temporal modeling of monthly forest fire occurrences was conducted using ARIMA, SARIMA, and SARIMAX approaches. Initially, ARIMA captured general trends but did not fully represent seasonal fire peaks. Introducing SARIMA improved alignment with seasonal patterns; however, its forecasting error remained higher compared to SARIMAX. The SARIMAX model incorporated relevant climatic predictors alongside AR and seasonal terms, allowing for both temporal and exogenous influences on fire dynamics.

The SARIMAX framework incorporates lagged temporal dependence through its AR, MA, and seasonal components. Consequently, historical fire occurrence patterns and seasonal persistence are represented within the model structure. Climatic predictors were included as temporally aligned monthly exogenous variables, while no additional manually specified climatic lag variables were introduced during model development.

All models were implemented in Python using the *statsmodels* and *pmdarima* libraries. Model performance was assessed using standard time-series metrics:

1) Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}$$

2) Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$

3) Mean Absolute Scaled Error (MASE)

$$MASE = \frac{\frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|}{\frac{1}{n}\sum_{j=1}^{m}\left|y_j - \hat{y}_{j-1}\right|}$$

An average MASE of 0.1902 indicates substantially lower forecasting error relative to a naïve seasonal baseline (MASE < 1). Based on its superior overall performance, SARIMAX was selected for further experiments (results are discussed in Section 4.3.1).

3.6.1. Exogenous predictor refinement for SARIMAX

Since ARIMA and SARIMA are univariate models, no external predictors are required. SARIMAX, however, incorporates exogenous climatic variables. Predictor refinement followed a structured four-step screening process to ensure relevance and reliability:

1) Removal of non-informative variables: SWE was excluded due to consistently zero values, while LC was omitted due to coarse annual resolution incompatible with monthly forecasting.
2) District-level correlation screening: Strongly correlated predictors with localized fire occurrence patterns were prioritized, reducing redundancy.
3) Multicollinearity filtering: Interdependence among predictors was assessed using VIF and TOL diagnostics. Predictors exceeding the predefined thresholds (VIF > 10 or TOL < 0.1) were removed to reduce redundancy, improve model stability, and prevent coefficient inflation.
4) Predictor combination testing: Multiple predictor configurations were evaluated to balance model parsimony with predictive performance. Dominant predictors varied by district, with drought severity indices (PDSI, KBDI) driving inland districts, while vegetation indices and runoff were important in coastal areas. Minimum temperature emerged as a key driver in several districts.

3.6.2. SARIMAX model specification and diagnostics

SARIMAX models were specified with a 12-month seasonal period. Model orders (p, d, q) and seasonal terms (P, D, Q) were determined using Akaike information criterion minimization within constrained ranges. Stationarity and absence of residual autocorrelation were verified, with differencing applied as needed. The inclusion of exogenous variables was controlled to prevent over-parameterization through prior multicollinearity screening and constrained parameter selection.

3.7. Spatial fire occurrence probability modeling

All classification models (GLM, MaxEnt, and ensemble) were evaluated using four key performance metrics: AUC, TSS, CBI, and accuracy.

AUC measures the model's ability to discriminate between positive and negative classes based on the ROC curve, which plots the true positive rate (TPR) against the false positive rate (FPR). An AUC value of 1 indicates perfect classification performance and is mathematically expressed as:

$$\boldsymbol{AUC} = \int_{0}^{1} \boldsymbol{TPR}\, \boldsymbol{d}\, (\boldsymbol{FPR}).$$

TSS evaluates classification performance while remaining insensitive to class imbalance and prevalence effects. It is computed as:

$$\boldsymbol{TSS = TPR - FPR},$$

where the TPR and FPR are defined as:

$$\boldsymbol{TPR} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FN}}$$

$$\boldsymbol{FPR} = \frac{\boldsymbol{FP}}{\boldsymbol{FP + TN}},$$

based on the confusion matrix components: true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

CBI evaluates the consistency between predicted suitability and the spatial distribution of observed occurrences. It assesses whether areas predicted as highly suitable contain a higher frequency of presences relative to random expectation. The index is computed using the Predicted-to-Expected (P/E) ratio across suitability classes and is mathematically expressed as the Spearman rank correlation between suitability ranks and P/E values:

$$\boldsymbol{CBI} = \boldsymbol{Spearman}\, \frac{\boldsymbol{P_i}}{\boldsymbol{E_i}},$$

where P_i and E_i represent the predicted and expected frequencies of occurrences within suitability class i. CBI values range from −1 to +1, with higher positive values indicating better model reliability.

Accuracy quantifies the overall proportion of correctly classified instances derived from the confusion matrix and is calculated as:

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP + TN}}{\boldsymbol{TP + TN + FP + FN}},$$

where TP and TN represent correctly predicted positive and negative cases, respectively, while FP and FN denote misclassifications. However, accuracy may be sensitive to class imbalance and is therefore interpreted alongside complementary performance metrics.

Together, these metrics provide a comprehensive assessment of model discrimination, probabilistic reliability, and classification performance.

3.7.1. Generalized Linear Model (GLM)

GLM applied a binomial logistic regression framework with fire occurrence (1 or 0) as the response variable. Climatic predictors passing multicollinearity screening were included. Regression coefficients and odds ratios were examined to interpret the magnitude and direction of predictor influence.

3.7.2. Maximum Entropy (MaxEnt)

MaxEnt estimated climatic suitability by contrasting fire presence locations with background conditions. L1-regularization controlled model complexity and prevented overfitting, while entropy maximization ensured the most uniform probability distribution under observed constraints. Permutation importance and response curves were analyzed to identify dominant climatic drivers and nonlinear thresholds.

3.7.3. Ensemble modeling

Predictions from GLM and MaxEnt were combined using AUC-weighted averaging. Individual model weights were derived from their respective AUC values, allowing models with stronger discrimination performance to contribute proportionally more to the final fire occurrence probability estimate. This ensemble structure balances GLM's interpretability with MaxEnt's nonlinear suitability estimation and was selected for further experiments based on superior overall performance (Section 4.4).

3.8. Hybrid SARIMAX–ensemble forecasting framework

To extend binary fire prediction into forward estimations, a hybrid framework was developed, integrating SARIMAX and ensemble models (Figure 8). For each district, predictors identified during refinement were projected over a 10-year horizon using SARIMAX-based climatic trajectories, with negative values clipped to zero. Historical data were used to train and evaluate GLM and MaxEnt models, whose predictions were then combined via ensemble averaging. Forecasted climatic variables from SARIMAX were input into the trained classifiers to estimate future fire probabilities, which were scaled using district-level historical mean fire occurrence rates to produce monthly fire count forecasts. Outputs were exported into a unified Comma-Separated Values (CSV) file, facilitating spatial and temporal fire-risk mapping for long-term planning and management.

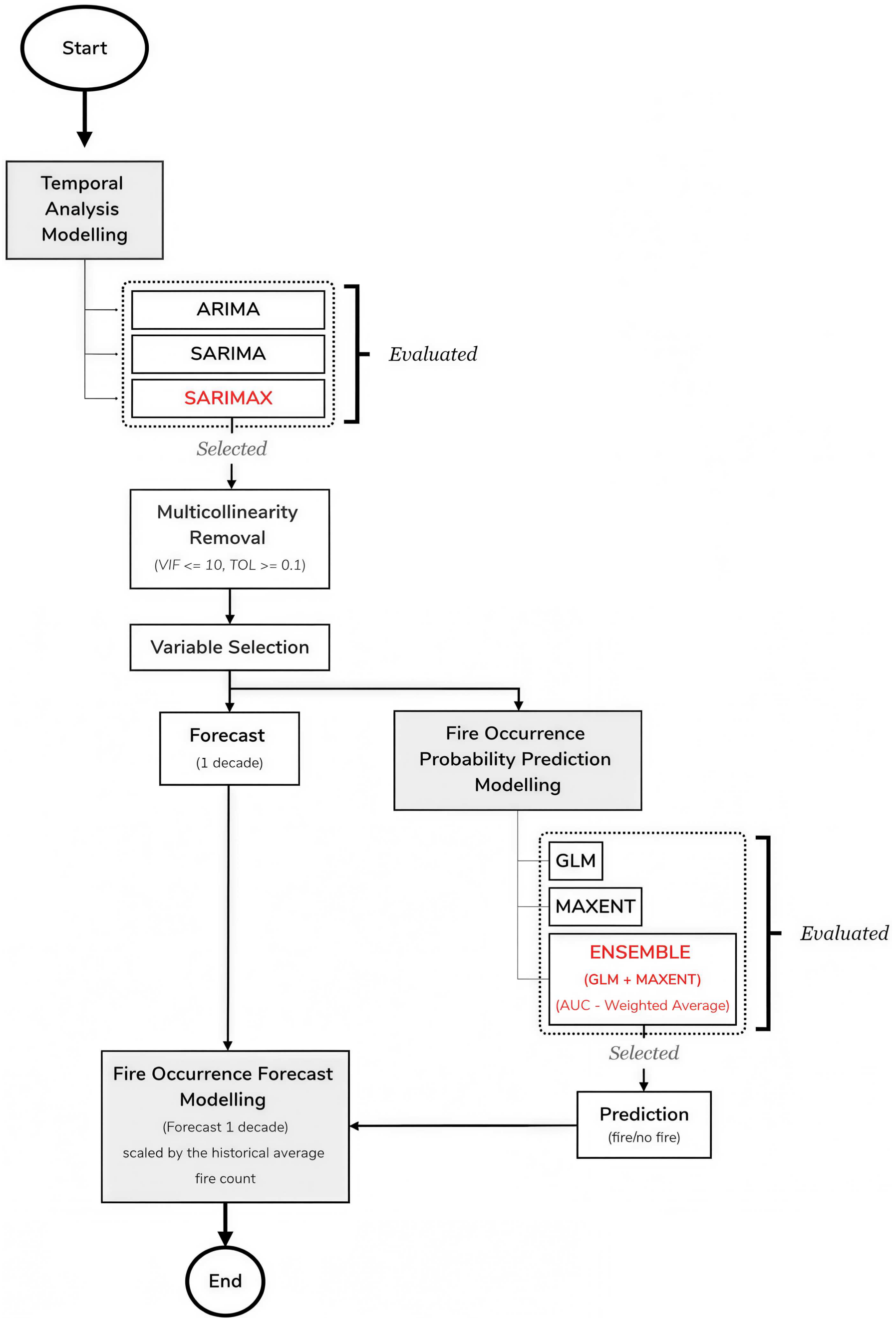
3.9. Dashboard visualization

All descriptive, temporal, probabilistic, and hybrid forecasting outputs were integrated into a Power BI dashboard. Python scripts were used for preprocessing, correlation analysis, SARIMAX forecasting, and ensemble predictions. The dashboard contains:

1) Fire trend charts
2) Climatic variable patterns
3) Correlation overlays
4) District-level probability maps
5) Forecasting panels

This platform enables interactive exploration of long-term fire trends and projected risk patterns, linking modeling outputs directly to actionable visual insights for stakeholders.

Figure 8
Hybrid SARIMAX–ensemble framework



4. Results

4.1. Descriptive and temporal patterns of fire occurrence

Between 2001 and 2023, forest fire occurrences across the 36 districts exhibited clear seasonal and interannual variability (Figure 9). Monthly aggregation showed a consistent concentration of fire activity during the first quarter of the year, with a pronounced peak in March. This seasonal pattern aligns with the northeast monsoon transition period, when rainfall declines and temperature rises, resulting in reduced surface and peat moisture.

Fire frequency increased notably during drought years, particularly 2005 and 2014. These periods correspond to intensified dry conditions associated with El Niño phases. Under such conditions, prolonged rainfall deficits reduced soil moisture while atmospheric evaporative demand intensified. As peat layers dried, ignition probability increased, and fires persisted for longer durations.

Spatially, several districts recorded recurrent fire activity (Figure 10). Coastal peat districts such as Sabak Bernam (Selangor) recorded the highest historical fire incidents (918), showing repeated seasonal spikes. Inland peat districts, including Pekan (Pahang), also recorded substantial fire counts (795), especially during extended dry periods.

Figure 9
Temporal distribution

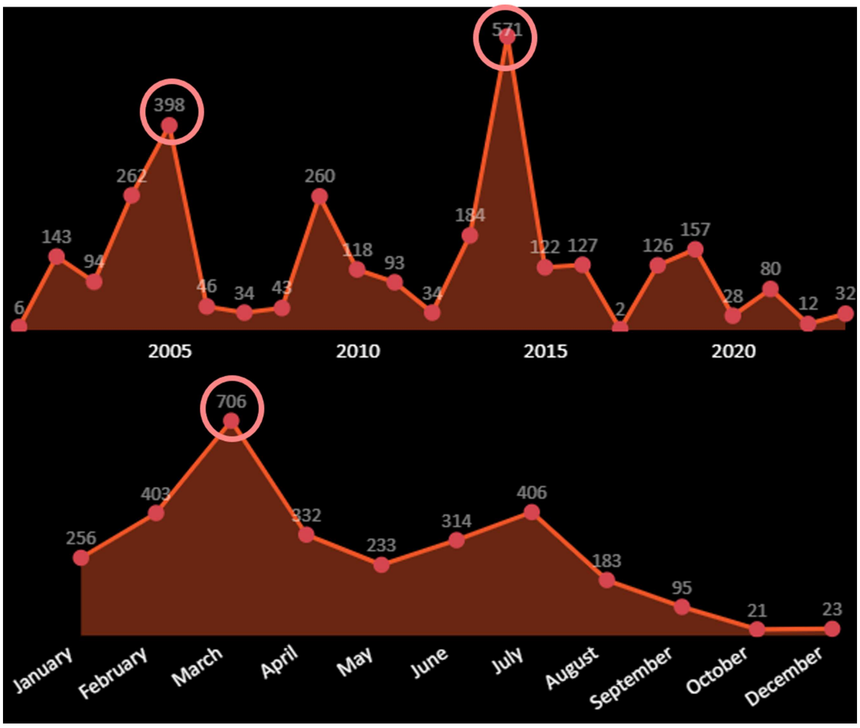
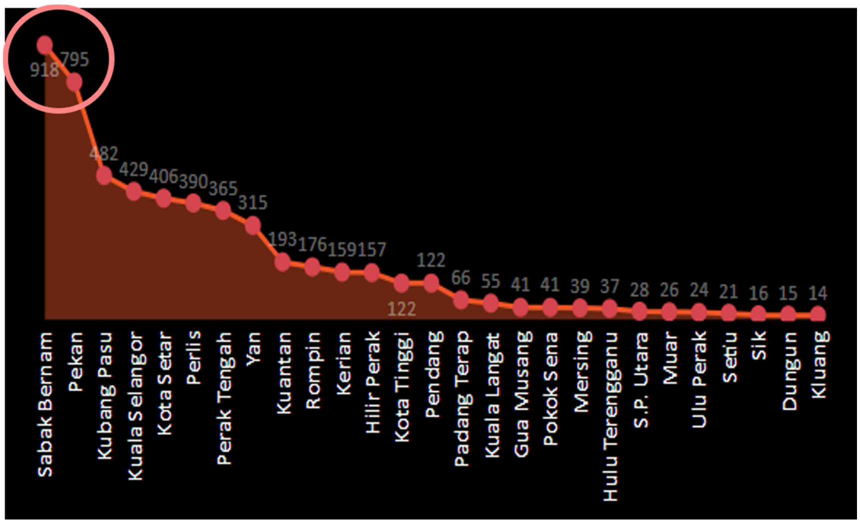


Figure 10
Spatial distribution



Emerging hotspots were detected outside traditional peat zones. Kubang Pasu, for instance, recorded the highest fire frequency (26 cases) in 2023 (Figure 11). This pattern suggests that fire risk is not confined solely to peat-dominated landscapes and may expand under intensified climatic stress.

Together, these observations suggest that seasonal climate cycles and interannual drought variability both influence fire occurrence patterns.

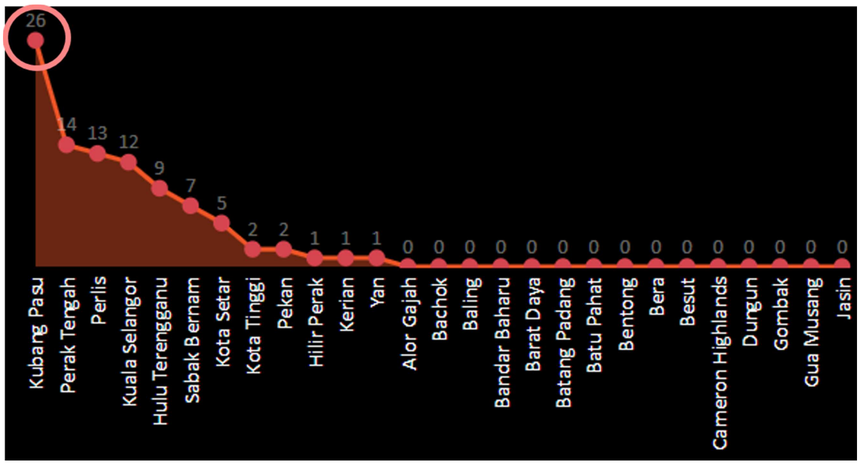
4.2. Climatic associations

Comparative analysis between fire and non-fire months revealed consistent climatic contrasts. Fire months were characterized by higher temperature values (including LST), higher drought severity (PDSI, KBDI), and increased water deficit. Conversely, fire events were linked to lower precipitation, reduced soil moisture, lower runoff, and decreased vegetation indices (NDVI, EVI).

These differences reflect progressive drying processes within peat swamp forests. Elevated temperatures intensify evapotranspiration, while reduced rainfall limits groundwater recharge. Declining NDVI and EVI values indicate vegetation stress and reduced canopy greenness preceding fire events.

Correlation analysis reinforced these patterns. Strong positive relationships were observed between fire occurrence and drought indices, whereas precipitation, runoff, soil moisture, NDVI, and

Figure 11
Hotspot detection in 2023



EVI showed negative associations. Such relationships support the theoretical premise that hydrological stress and vegetation dryness increase flammability.

District-level variation was also evident. In several inland districts (e.g., Pahang and Terengganu), PDSI and climate water deficit were the predominant drivers. In contrast, coastal districts like Sepang and Klang displayed closer associations with minimum temperature and NDVI. This spatial variability supports the use of district-specific predictor selection in SARIMAX modeling.

4.3. Temporal modeling performance

4.3.1. Model comparison

ARIMA, SARIMA, and SARIMAX models were evaluated using RMSE, MAE, and MASE (Table 1). ARIMA captured the general trend direction but did not adequately reproduce the recurring March peak. Incorporating 12-month seasonal terms improved alignment in SARIMA, yet forecast errors remained higher than those of the SARIMAX model.

Table 1
Temporal models performance summary

Model	RMSE	MAE	MASE
ARIMA	0.02	0.02	0.77
SARIMA	0.35	0.19	0.22
SARIMAX	0.02	0.00	0.19

After incorporating exogenous climatic predictors, SARIMAX showed consistently stronger performance than both baseline models across the 36 districts. The average MASE for SARIMAX was 0.1902. A MASE value below 1 indicates performance superior to a naïve seasonal baseline. A value of 0.1902 implies that the average absolute forecast error is approximately 19% of the error produced by a simple seasonal benchmark. In practical terms, this means SARIMAX reduces forecast error by about 81% compared with simply repeating the previous year's monthly fire count. This reflects strong predictive capability in capturing climate-driven fire variability.

Lower RMSE (0.0159) and MAE (0.0036) values further confirm improved forecast precision. Reduced RMSE indicates fewer large deviations, while lower MAE reflects improved average absolute accuracy.

4.3.2. District-level drivers

The SARIMAX variable refinement process identified distinct dominant predictors across districts (Table 2).

1) Inland peat districts were strongly influenced by drought severity indices (PDSI, KBDI) and water deficit.
2) Coastal districts showed stronger relationships with vegetation health metrics (NDVI, EVI) and hydrological factors (e.g., runoff).
3) In several districts, minimum temperature emerged as the primary driver.

These results indicate that while drought acts as a consistent regional driver, local climatic sensitivity varies spatially.

4.3.3. Residual diagnostics

Following model fitting, residual patterns were examined as part of model diagnostics to assess whether observable temporal structure remained after estimation. Residuals generally fluctuated around zero without evident recurring seasonal patterns, suggesting that model specification and differencing reduced observable temporal dependence. These findings are interpreted as diagnostic observations based on residual inspection and should not be considered formal statistical verification of residual independence or absence of autocorrelation. Stationarity was supported through differencing procedures applied prior to estimation. Overall, residual inspection suggested that major temporal patterns were adequately represented within the SARIMAX framework.

4.4. Spatial fire probability modeling

4.4.1. GLM results

The GLM logistic regression model achieved an AUC of 1 alongside perfect TSS and accuracy values (Table 3). This indicates near-perfect discrimination between fire and non-fire months at the district–monthly aggregation scale under the defined binary threshold. The strong separability likely reflects pronounced climatic contrasts within aggregated data and

Table 2
Optimized variable combinations in peat swamp forest states

State	District	Best combination of variables	RMSE	MAE	MASE	Variables selected
Pahang	Pekan	17	0.0161	0.0047	0.1403	[“def,” “pdsi”]
	Kuantan	13 (exclude: vpd, pdsi, soil, def)	0.0000	0.0000	0.0000	[“ro”]
	Rompin	16 (exclude: pdsi)	$5.22e^{-5}$	$1.17e^{-5}$	0.0003	[“ro,” “def”]
	Bentong	17	0.0000	0.0000	0.0000	[“soil,” “pdsi”]
	Lipis	17	0.0000	0.0000	0.0000	[“pdsi,” “def”]
	Maran	17	0.0000	0.0000	0.0000	[“KBDI,” “pdsi”]
	Raub	17	$1.71e^{-11}$	$3.49e^{-12}$	$1.95e^{-10}$	[“ro,” “def”]
	Temerloh	17	0.0000	0.0000	0.0000	[“def,” “KBDI,” “pdsi”]
	Bera	17	0.0000	0.0000	0.0000	[“pdsi,” “ro,” “def”]
	Jerantut	17	0.0000	0.0000	0.0000	[“ro,” “def”]
	Cameron Highlands	17	0.0000	0.0000	0.0000	[“def”]
Selangor	Kuala Selangor	15 (exclude: def, soil)	0.0265	0.0071	0.1224	[“pdsi,” “KBDI”]
	Sabak Bernam	17	0.0262	0.0061	0.0805	[“pdsi,” “soil”]
	Klang	14 (exclude: pr, pdsi, ro)	0.0005	0.0001	0.0064	[“def,” “NDVI”]
	Kuala Langat	16 (exclude: pdsi)	0.0014	0.0004	0.0146	[“ro,” “def”]
	Sepang	10(exclude: KBDI, NDVI, vs, EVI, soil, def, pdsi)	$9.32e^{-136}$	$1.97e^{-136}$	$1.47e^{-134}$	[“tmnn”]
	Ulu Selangor	13 (exclude: vap, pdsi, vs, def)	0.1284	0.0262	2.2409	[“NDVI”]
	Ulu Langat	–	–	–	–	–
	Gombak	–	–	–	–	–
Johor	Segamat	17	0.0000	0.0000	0.0000	[“ro,” “pdsi”]
	Kluang	17	0.0011	0.0004	0.0262	[“def,” “vpd,” “ro”]
	Muar	17	0.0000	0.0000	0.0000	[“pdsi,” “def”]
	Batu Pahat	13 (exclude: vpd, vs, pdsi, def)	0.0015	0.0004	0.0322	[“soil”]
	Kota Tinggi	11 (exclude: NDVI, EVI, pr, def, ro, pdsi)	0.0295	0.0091	0.3244	[“KBDI”]
	Mersing	17	0.0000	0.0000	0.0000	[“pdsi,” “def”]
	Johor Bahru	17	0.0000	0.0000	0.0000	[“pet”]
	Pontian	17	0.0000	0.0000	0.0000	[“pdsi,” “def”]
	Kulaijaya	17	0.0000	0.0000	0.0000	[“vap”]
	Ledang	17	0.0000	0.0000	0.0000	[“ro”]
Terengganu	Dungun	16 (exclude: def)	$3.60e^{-5}$	$3.60e^{-5}$	0.0015	[“soil,” “pdsi”]
	Besut	16 (exclude: pdsi)	$3.65e^{-12}$	$1.73e^{-12}$	$1.23e^{-10}$	[“NDVI,” “soil”]

(*Continued*)

Table 2
(*Continued*)

State	District	Best combination of variables	RMSE	MAE	MASE	Variables selected
	Kemaman	13 (exclude: vpd, soil, pdsi, def)	0.0000	0.0000	0.0000	[“ro”]
	Kuala Terengganu	17	$2.82e^{-6}$	$2.82e^{-6}$	0.0001	[“ro,” “pdsi,” “def”]
	Marang	17	$6.20e^{-45}$	$1.27e^{-45}$	$9.54e^{-44}$	[“soil,” “pdsi”]
	Setiu	13 (exclude: soil, ro, pdsi, def)	0.0000	0.0000	0.0000	[“KBDI”]
	Hulu Terengganu	17	0.0745	0.0214	1.1182	[“pdsi,” “ro,” “def”]

Note: Several districts produced near-zero RMSE, MAE, and MASE values. These values do not indicate perfect predictive performance and should be interpreted alongside the discussion in Section 4.5, where the influence of low temporal variability and prolonged zero-fire periods is described.

should be interpreted within the simplified classification structure (fire ≥ 1 incident per month).

Regression coefficients showed that higher drought severity, elevated temperature, and greater water deficit increased fire probability. Conversely, higher precipitation, soil moisture, runoff, NDVI, and EVI reduced fire likelihood. Odds ratios further indicated substantial marginal influence from drought indices.

4.4.2. MaxEnt results

MaxEnt also demonstrated strong predictive performance with an AUC of 1, though it yielded a lower TSS (0.3057) and CBI (0.5095) compared to GLM (Table 3). Response curves revealed nonlinear relationships between climatic predictors and fire probability. Suitability increased sharply beyond specific drought thresholds, reflecting critical moisture limits within peat systems.

Permutation importance analysis confirmed the dominant role of drought indices and temperature variables.

4.4.3. Ensemble performance

Predictions from GLM and MaxEnt were combined using AUC-weighted averaging. Ensemble performance achieved the highest CBI of 0.67 among the classification models (Table 3), indicating moderate to strong reliability in ranking climatic suitability across the 36 districts.

This integrated approach combines GLM’s coefficient interpretability with MaxEnt’s nonlinear flexibility, improving the stability of probability estimates and overall predictive accuracy.

4.5. Hybrid forecasting results

The hybrid SARIMAX–ensemble framework extended binary fire classification into forward-looking projections by forecasting district-level climatic predictors over a 10-year horizon using SARIMAX, with negative values clipped to zero to preserve physical plausibility. The near-zero error values observed in several districts do not indicate perfect predictive performance. Instead, they arise from extremely low temporal variability in the fire occurrence series, where prolonged zero-fire sequences dominate both the training and validation periods. Under such zero-inflated conditions, model forecasts converge to persistent zero predictions, producing RMSE, MAE, and MASE values of 0.000 when evaluated against equally zero observations. This reflects data sparsity rather than unrealistically precise model accuracy.

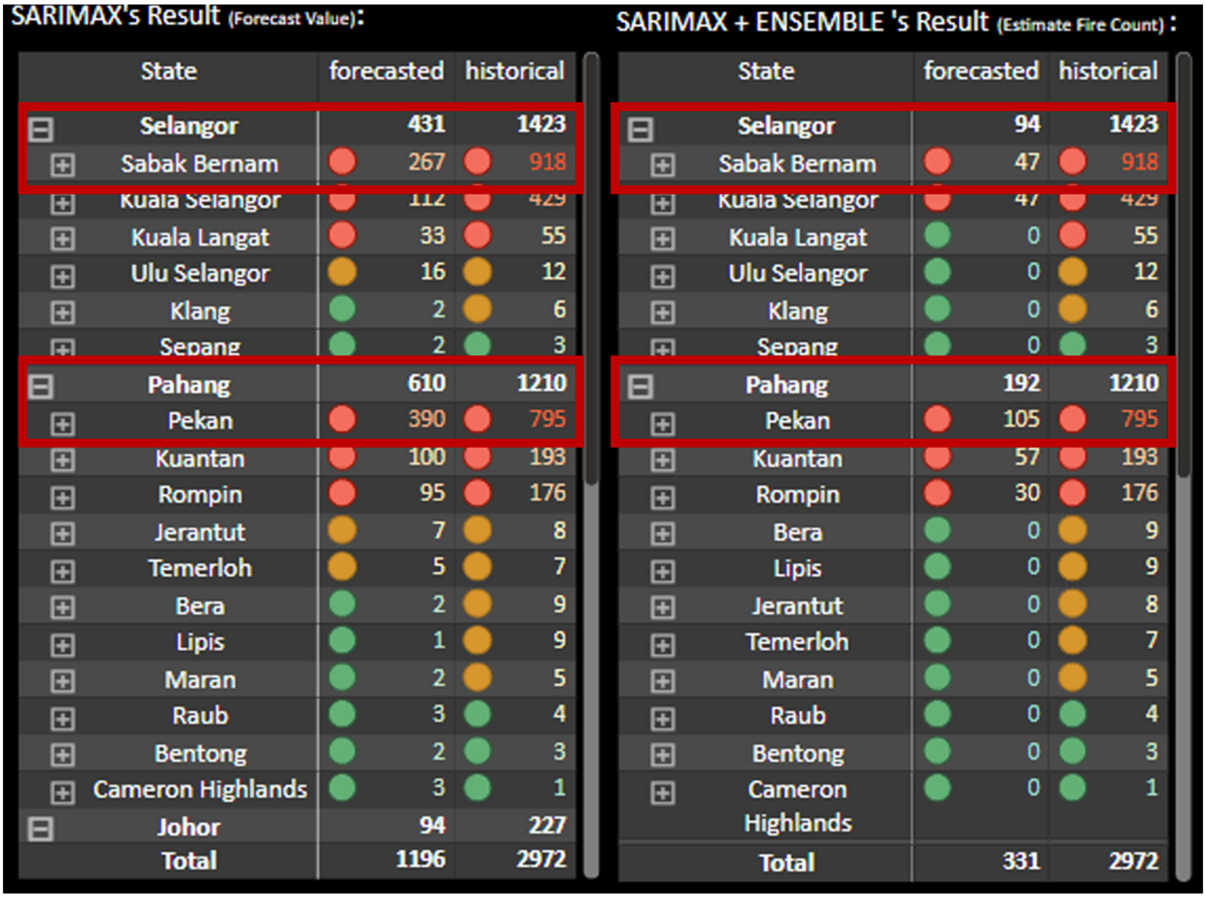
These projected predictors were then input into the trained GLM and MaxEnt classifiers. Ensemble probability outputs were subsequently scaled using the historical mean fire count observed during fire occurrence periods to estimate projected monthly fire counts.

Projection outputs suggest a gradual spatial redistribution of fire risk across Peninsular Malaysia. Coastal hotspot concentration in Sabak Bernam showed a moderate decline relative to inland peat districts such as Pekan, indicating a potential shift in fire activity over the next decade. SARIMAX projected approximately 431 cumulative cases in Selangor and 610 cases in Pahang over the projection period, whereas ensemble projections produced comparatively lower estimates (94 and 192 cases, respectively), as illustrated in Figure 12. Differences between SARIMAX and ensemble outputs reflect methodological distinctions between direct temporal extrapolation and probability-based

Table 3
Classification models performance summary

Model	AUC	TSS	CBI	Accuracy
GLM	1	1	0.52	1
MaxEnt	1	0.31	0.51	0.98
Ensemble Models	1	1	0.67	1

Figure 12
Fire hotspot shift



SARIMAX's Result (Forecast Value):

State	forecasted	historical
Selangor	431	1423
Sabak Bernam	267	918
Kuala Selangor	112	429
Kuala Langat	33	55
Ulu Selangor	16	12
Klang	2	6
Sepang	2	3
Pahang	610	1210
Pekan	390	795
Kuantan	100	193
Rompin	95	176
Jerantut	7	8
Temerloh	5	7
Bera	2	9
Lipis	1	9
Maran	2	5
Raub	3	4
Bentong	2	3
Cameron Highlands	3	1
Johor	94	227
Total	1196	2972

SARIMAX + ENSEMBLE 's Result (Estimate Fire Count) :

State	forecasted	historical
Selangor	94	1423
Sabak Bernam	47	918
Kuala Selangor	47	429
Kuala Langat	0	55
Ulu Selangor	0	12
Klang	0	6
Sepang	0	3
Pahang	192	1210
Pekan	105	795
Kuantan	57	193
Rompin	30	176
Bera	0	9
Lipis	0	9
Jerantut	0	8
Temerloh	0	7
Maran	0	5
Raub	0	4
Bentong	0	3
Cameron Highlands	0	1
Total	331	2972

scaling. These projections represent scenario-based estimates derived from projected climatic conditions and should be interpreted as indicative patterns of future fire distribution rather than probabilistic or confidence-bounded forecasts.

Seasonal March peaks persist in projected outputs, indicating structural seasonality rather than isolated anomalies. The hybrid framework thus translates climatic forecasts into operationally interpretable fire count projections.

5. Discussion

5.1. Contribution to literature

The findings help address several gaps identified in earlier forest fire research conducted in Peninsular Malaysia.

First, the 22-year dataset captures multiple ENSO cycles. Shorter datasets (<10 years) may detect isolated drought events but cannot robustly distinguish structural seasonal patterns from episodic anomalies [5, 14]. The extended temporal span allows models to learn recurring March peaks alongside repeated drought responses.

Second, the analysis integrates hydrological and vegetation stress indicators in addition to rainfall and temperature [14]. This integration is particularly relevant in peat swamp ecosystems, where groundwater decline directly increases flammability [3]. El Niño events are fundamentally associated with reduced rainfall and elevated regional temperatures [3], conditions that lower groundwater recharge while increasing evapotranspiration, thereby accelerating peat drying [6]. Drier peat reduces ignition thresholds and supports longer fire persistence [4], often contributing to severe haze episodes [5]. Observed fire peaks in 2005 and 2014 coincide with intensified drought conditions, reinforcing the climatic mechanism linking ENSO variability, peat hydrology, and fire susceptibility.

Third, the study combines temporal forecasting with probabilistic modeling within a hybrid framework. Earlier studies often applied these approaches independently [11, 12]. Linking SARIMAX climatic projections with ensemble classifiers enables forward-looking fire probability and count estimation within a unified system.

5.2. Policy and management implications

The modeling framework presents several operational implications for fire management in Peninsular Malaysia:

1) Early Warning Systems: Seasonal peaks in March enable agencies to prioritize monitoring and resource allocation ahead of critical periods.
2) Drought Monitoring Integration: Strong links with PDSI and KBDI indicate that drought indices can function as early fire-risk indicators.
3) Spatial Targeting: District-specific predictors support tailored mitigation strategies rather than uniform national approaches.
4) Long-Term Planning: Hybrid projections offer a 10-year planning horizon for peatland management and groundwater regulation.

By converting climatic variability into measurable fire-risk indicators, the framework supports a more proactive management approach rather than a purely reactive one.

5.3. Limitations

Several limitations should be acknowledged.

1) Temporal aggregation: Monthly aggregation may smooth short-duration fire events and extreme daily spikes.
2) Spatial aggregation: District-level aggregation may mask sub-district land-use and ignition variability.
3) Binary threshold definition: Defining fire occurrence as ≥1 incident per month may simplify classification and contribute to very high AUC values.
4) Missing value removal: Excluding persistently missing variables improves stability but may introduce bias if missingness was systematic (e.g., sensor interruption during extreme years).
5) Model assumptions: SARIMAX assumes linear predictor–response relationships, which may not capture nonlinear ecological thresholds.
6) Uncertainty quantification: Forecast outputs were generated as point estimates using the fitted SARIMAX forecasting framework and therefore represent indicative future scenarios derived from projected climatic trajectories. Uncertainty ranges were not estimated within the current implementation and remain an opportunity for future extension.
7) Scope of study: The study's current scope excludes Sabah and Sarawak, limiting its application to the entire Malaysian context.
8) Technical and accessibility issues: The Power BI dashboard's Python visualizations could not be rendered in web mode without a Pro License, limiting accessibility and broader dissemination of interactive results.

Despite these constraints, the integrated framework remains robust for long-term climate-driven fire assessment.

6. Conclusion

The study presents a 22-year assessment of climatic–forest fire relationships within peat swamp regions of Peninsular Malaysia. Long-term analysis confirms that hydroclimatic variability, particularly the ENSO-driven warming trend of approximately 0.25 °C per decade, plays a dominant role in shaping seasonal and interannual fire patterns.

SARIMAX achieved the strongest temporal forecasting performance (MASE = 0.1902), demonstrating substantial (around 81%) improvement over naïve seasonal baselines. GLM and MaxEnt showed strong climatic separability in predicting fire occurrence, while ensemble averaging improved the reliability of suitability rankings with a CBI of 0.67.

The hybrid SARIMAX–ensemble framework extends temporal forecasting into spatially explicit probability and fire count projections. Findings indicate a gradual redistribution of hotspot intensity, with a projected shift from coastal zones like Sabak Bernam toward inland peat districts like Pekan under continued drought variability.

Overall, the integrated modeling system supports climate-informed fire monitoring while contributing to early warning strategies and long-term peatland management planning.

7. Recommendations and Future Direction

The hybrid SARIMAX–ensemble framework provides a structured basis for operational integration into state-level fire monitoring systems. Automated updating of climatic inputs and scheduled model recalibration would enable dynamic adjustment to evolving drought conditions, particularly by monitoring fluctuations in the KBDI and PDSI [17, 18]. Translating probability outputs into categorical risk levels may further improve communication with district enforcement units, facilitating faster response times during peak periods like March.

The 10-year projection capability offers a medium-term planning horizon. Future frameworks may incorporate projected fire probability into peatland water management strategies to mitigate the high flammability of organic soils [3]. For inland peat districts identified as increasingly vulnerable, such as Pekan, groundwater stabilization should be prioritized within long-term land management plans to prevent the water table from dropping below critical thresholds [3]. Integrating projected fire risk into spatial zoning decisions may also reduce exposure of highly flammable peat areas. Instead of reacting to annual outbreaks, authorities may use projected seasonal concentration patterns to plan equipment allocation, training schedules, and monitoring expansion ahead of time.

Future studies should strengthen statistical robustness in four areas. First, formal residual diagnostic testing, including Ljung–Box tests, additional autocorrelation checks, validation against observed fire occurrence records, cross-validation, and comparison with actual fire occurrences, may further validate temporal independence assumptions [10, 11, 22–25]. Second, incorporating prediction intervals for SARIMAX forecasts [18] would enable quantification of uncertainty ranges rather than reliance on point estimates alone, addressing a current limitation in fire frequency modeling. Third, testing alternative binary thresholds for fire occurrence would assess classification stability under stricter event definitions, ensuring the models remain accurate as ignition patterns evolve. The extremely high classification performance observed in district–monthly aggregated data may also reflect reduced variability introduced through temporal aggregation and simplified event definition. Fourth, future work may incorporate interpretable machine-learning techniques, such as SHapley Additive exPlanations, to quantify the marginal contribution of climatic predictors and reveal interaction effects underlying fire-risk predictions, thereby complementing predictive evaluation with clearer ecological interpretation.

To support reproducible research, the processed district–monthly dataset should be accompanied by a structured data dictionary describing variable units and definitions, building upon the open-source framework established by Chew et al. [17]. Continued data expansion beyond 2023 will improve recalibration of long-term models and strengthen predictive stability.

The hybrid approach linking SARIMAX forecasting with ensemble classification (GLM and MaxEnt) may also be extended to include socioeconomic and land-use data to better capture anthropogenic influences [1, 12, 16]. The present study focused exclusively on climatic and environmental variables available within the forest fire inventory dataset; therefore, factors such as agricultural burning, land-use change, and slash-and-burn activities were beyond the scope of the current analysis. Future integration of these anthropogenic variables may further improve understanding of fire occurrence dynamics and enhance predictive performance. Climatic scenario inputs could evaluate the sensitivity of projected fire risk under intensified drought conditions or extreme ENSO phases. Cross-regional application in other peat-dominated landscapes would allow assessment of transferability and structural robustness without altering the current methodology.

Acknowledgment

The authors acknowledge the use of the forest fire dataset from Chew et al. [17]. Development of the analytical components benefited from the integration of time-series forecasting and spatial modeling tools, alongside visualization functions implemented through the Power BI environment. The computational infrastructure and datasets enabled long-term examination of climatic–fire relationships across the 22-year study period (2001–2023) in Peninsular Malaysia.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Zenodo at https://zenodo.org/records/11542164.

Author Contribution Statement

Grace Kian Hwai Ling: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Weng Howe Chan:** Conceptualization, Validation, Writing – review & editing, Supervision, Project administration.

References

[1] Celis, N., Casallas, A., Lopez-Barrera, E. A., Felician, M., de Marchi, M., & Pappalardo, S. E. (2023). Climate change, forest fires, and territorial dynamics in the Amazon rainforest: An integrated analysis for mitigation strategies. *ISPRS International Journal of Geo-Information*, *12*(10), 436. https://doi.org/10.3390/ijgi12100436

[2] Gajendiran, K., Kandasamy, S., & Narayanan, M. (2024). Influences of wildfire on the forest ecosystem and climate change: A comprehensive study. *Environmental Research*, *240*, 117537. https://doi.org/10.1016/j.envres.2023.117537

[3] Li, L., Sali, A., Noordin, N. K., Ismail, A., & Hashim, F. (2023). Prediction of peatlands forest fires in Malaysia using machine learning. *Forests*, *14*(7), 1472. https://doi.org/10.3390/f14071472

[4] Ahmad, W. M. S. W. (n.d.). *Fire situation in Malaysia*. Food and Agriculture Organization of the United Nations. https://www.fao.org/4/ad653e/ad653e46.htm#P3907_325042

[5] Chew, Y. J., Ooi, S. Y., Pang, Y. H., & Wong, K.-S. (2022). A review of forest fire combating efforts, challenges and future directions in Peninsular Malaysia, Sabah, and Sarawak. *Forests*, *13*(9), 1405. https://doi.org/10.3390/f13091405

[6] Santika, T., Budiharta, S., Law, E. A., Dennis, R. A., Dohong, A., Struebig, M. J., . . . , & Wilson, K. A. (2020). Interannual climate variation, land type and village livelihood effects on fires in Kalimantan, Indonesia. *Global Environmental Change*, *64*, 102129. https://doi.org/10.1016/j.gloenvcha.2020.102129

[7] Jones, M. W., Abatzoglou, J. T., Veraverbeke, S., Andela, N., Lasslop, G., Forkel, M., . . . , & le Quéré, C. (2022). Global and regional trends and drivers of fire under climate change. *Reviews of Geophysics*, *60*(3), e2020RG000726. https://doi.org/10.1029/2020RG000726

[8] Burton, C., Lampe, S., Kelley, D. I., Thiery, W., Hantson, S., Christidis, N., . . . , & Mengel, M. (2024). Global burned area increasingly explained by climate change. *Nature Climate Change*, *14*(11), 1186–1192. https://doi.org/10.1038/s41558-024-02140-w

[9] Iskandar Shah, D. R. S. (2023). *Climate changes in Malaysia: Wake up! It's real*. Azman Hashim International Business School. https://business.utm.my/blog/2023/12/18/climate-changes-in-malaysia-wake-up-its-real/

[10] Kale, M. P., Mishra, A., Pardeshi, S., Ghosh, S., Pai, D. S., & Roy, P. S. (2022). Forecasting wildfires in major forest types of India. *Frontiers in Forests and Global Change*, *5*, 882685. https://doi.org/10.3389/ffgc.2022.882685

[11] Kouassi, J.-L., Wandan, N., & Mbow, C. (2020). Predictive modeling of wildfire occurrence and damage in a tropical savanna ecosystem of West Africa. *Fire*, *3*(3), 42. https://doi.org/10.3390/fire3030042

[12] Babu, K. N., Gour, R., Ayushi, K., Ayyappan, N., & Parthasarathy, N. (2023). Environmental drivers and spatial prediction of forest fires in the Western Ghats biodiversity hotspot, India: An ensemble machine learning approach. *Forest Ecology and Management*, *540*, 121057. https://doi.org/10.1016/j.foreco.2023.121057

[13] Prasetyo, L. B., Setiawan, Y., Condro, A. A., Kustiyo, K., Putra, E. I., Hayati, N., . . . , & Murdiyarso, D. (2022). Assessing Sumatran peat vulnerability to fire under various condition of ENSO phases using machine learning approaches. *Forests*, *13*(6), 828. https://doi.org/10.3390/f13060828

[14] Chew, Y. J., Ooi, S. Y., Pang, Y. H., & Lim, Z. Y. (2023). Investigating the relationship between the influencing fire factors and forest fire occurrence in the districts of Rompin, Pekan, and Kuantan in the state of Pahang, Malaysia, using Google Earth Engine. *International Journal on Advanced Science, Engineering and Information Technology*, *13*(5), 1733–1741. https://doi.org/10.18517/ijaseit.13.5.19026

[15] Abram, N. J., Henley, B. J., Sen Gupta, A., Lippmann, T. J. R., Clarke, H., Dowdy, A. J., . . . , & Boer, M. M. (2021). Connections of climate change and variability to large and extreme forest fires in southeast Australia. *Communications Earth & Environment*, *2*(1), 8. https://doi.org/10.1038/s43247-020-00065-8

[16] Vilar, L., Herrera, S., Tafur-García, E., Yebra, M., Martínez-Vega, J., Echavarría, P., & Martín, M. P. (2021). Modelling wildfire occurrence at regional scale from land use/cover and climate change scenarios. *Environmental Modelling & Software*, *145*, 105200. https://doi.org/10.1016/j.envsoft.2021.105200

[17] Chew, Y. J., Ooi, S. Y., Pang, Y. H., & Lim, Z. Y. (2024). Framework to create inventory dataset for disaster behavior analysis using Google Earth Engine: A case study in Peninsular Malaysia for historical forest fire behavior analysis. *Forests*, *15*(6), 923. https://doi.org/10.3390/f15060923

[18] Orero, L., Omondi, E. O., & Omolo, B. O. (2024). A Bayesian model for predicting monthly fire frequency in Kenya. *PLOS ONE*, *19*(1), e0291800. https://doi.org/10.1371/journal.pone.0291800

[19] Joshi, K. P., Adhikari, G., Bhattarai, D., Adhikari, A., & Lamichanne, S. (2024). Forest fire vulnerability in Nepal's Chure region: Investigating the influencing factors using generalized linear model. *Heliyon*, *10*(7), e28525. https://doi.org/10.1016/j.heliyon.2024.e28525

[20] Trang, P. T., Andrew, M. E., Chu, T., & Enright, N. J. (2022). Forest fire and its key drivers in the tropical forests of northern Vietnam. *International Journal of Wildland Fire*, *31*(3), 213–229. https://doi.org/10.1071/WF21078

[21] United Nations Development Programme. (2006). *Malaysia's peat swamp forests: Conventional and sustainable use*. https://files.acquia.undp.org/public/migration/my/malaysia-peat-swamp-forest.pdf

[22] Ardiyani, E., Nurdiati, S., Sopaheluwakan, A., Septiawan, P., & Najib, M. K. (2023). Probabilistic hotspot prediction model based on Bayesian inference using precipitation, relative dry spells, ENSO and IOD. *Atmosphere*, *14*(2), 286. https://doi.org/10.3390/atmos14020286

[23] Dhar, T., Bhatta, B., & Aravindan, S. (2023). Forest fire occurrence, distribution and risk mapping using geoinformation technology: A case study in the sub-tropical forest of the Meghalaya, India. *Remote Sensing Applications: Society and Environment*, *29*, 100883. https://doi.org/10.1016/j.rsase.2022.100883

[24] Ribeiro, A. F. S., Brando, P. M., Santos, L., Rattis, L., Hirschi, M., Hauser, M., . . . , & Zscheischler, J. (2022). A compound event-oriented framework to tropical fire risk assessment in a changing climate. *Environmental Research Letters*, *17*(6), 065015. https://doi.org/10.1088/1748-9326/ac7342

[25] Williamson, G. J., Tng, D. Y., & Bowman, D. M. (2024). Climate, fire, and anthropogenic disturbance determine the current global distribution of tropical forest and savanna. *Environmental Research Letters*, *19*(2), 024032. https://doi.org/10.1088/1748-9326/ad20ac

How to Cite: Ling, G. K. H., & Chan, W. H. (2026). Spatiotemporal Analysis of Climate–Forest Fire Correlation in Peninsular Malaysia Using SARIMAX and Ensemble Models. *Journal of Data Science and Intelligent Systems*. https://doi.org/10.47852/bonviewJDSIS62028499