IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 November 2025, accepted 6 December 2025, date of publication 11 December 2025,
date of current version 18 December 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3643170

RESEARCH ARTICLE

Assessing Fire Risk Homogeneity and Patterns With GIS and Machine Learning: A Study From the Czech Republic

DAVID SLÁDEK[1], FILIP DOHNAL[2], FRANTIŠEK PAULUS[3], JIŘÍ NEUBAUER[4], MARTIN MOLEK[3], AND TOMÁŠ ZEMAN[2]

[1]Department of Military Geography and Meteorology, Faculty of Military Technology, University of Defence, 66210 Brno, Czech Republic
[2]Department of Societal Security, Faculty of Logistics and Crisis Management, Tomas Bata University in Zlín, 68601 Uherské Hradiště, Czech Republic
[3]Population Protection Institute, 53341 Lázně Bohdaneč, Czech Republic
[4]Department of Quantitative Methods, Faculty of Military Leadership, University of Defence, 66210 Brno, Czech Republic

Corresponding author: Filip Dohnal (fdohnal@utb.cz)

This work was supported in part by the Internal Project Assessment of Territorial Vulnerability to Current Security Threats at the Tomas Bata University in Zlín under Grant RVO/FLKŘ/2024/02; in part by the Project Creativity, Intelligence and Talent pro Zlínský kraj; and in part by the Internal Project Military Autonomous and Robotic Assets at the University of Defence in Brno under Grant DZRO-FVT22-VAROPS.

ABSTRACT Forest fires have historically been a concern in fire-prone regions, where climatic and environmental conditions naturally foster fire occurrences. However, in Central Europe, the increasing frequency of wildfires is driven by a complex interplay of factors, reflecting both environmental changes and human influences. This study uses machine learning to address key challenges in understanding fire risk in low-risk regions, specifically: the limited applicability of models developed for high-risk areas and the struggle to incorporate dynamic spatiotemporal factors. By analyzing eight years of fire data alongside variables such as weather, geography, and population density, our objective is to develop a machine learning and GIS-based procedure for identifying and understanding the underlying complexity of fire risk in Czechia. K-Medoids clustering revealed challenges in forming distinct clusters, suggesting the need for additional or more refined predictors. Classification models (e.g. Random Forest, Gradient Boosting, k-Nearest Neighbors) were used to predict periods of increased fire risk. Soil temperature and population density emerged as the most significant predictors. This study proposed two procedures: one that combines clustering and bivariate analysis to identify clusters with homogeneous factors, and another as a classification-bivariate visualization framework for assessing fire risk. The results indicate that classification performance can serve as a proxy for spatial complexity and highlight regions where multiple factors drive fire occurrence. To enhance model accuracy and applicability, incorporating high-resolution weather data, population mobility patterns, vegetation types, and advanced modeling techniques is recommended. Introducing advanced predictors could help differentiate between fires driven by natural factors and those influenced by human activities.

INDEX TERMS Wildfire, forest fire, fire risk, machine learning, clustering, GIS, spatial analysis, GFS model.

I. INTRODUCTION

Open fires, encompassing wildland and grass fires, pose substantial risks to ecosystems, economies, and human health worldwide. Predicting their occurrence is a critical step toward mitigating these risks and implementing effective prevention strategies. In this study from the area of the Czech Republic, "open fire" is defined as a fire classified by firefighters as either wildland or grass fire. Our research focuses on integrating Geographic Information Systems (GIS),

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tong.

TABLE 1. Forest fire drivers and prediction literature overview.

Authors	Year	Area	Main Conclusion
Huot et al. [1]	2022	Continental US	Data with resolution of 1 km are too coarse to capture fine-scale details, such as local wind variations and the impact of human activity.
Mambile et al [3]	2024	Global	Remote sensing data show promise, but lack of human activity data and over-reliance on simulations instead of real-world results limit the progress.
Muhs et al. [4]	2020	Unspecified	Quantifying wildfire risk and the effectiveness of mitigation techniques is a foremost challenge, requiring the development of unified, probabilistic models based on historical data.
Chuevico et al. [5]	2012	Peninsular Spain	Fire risk assessment system for modelling drivers (including human factors and fuel conditions) and fire vulnerability using high-resolution spatial data should use GIS capabilities.
Ganteaume et al. [6]	2013	Mediterranean	Human and environmental factors influence fire ignition, with regional variations requiring specific management approaches.
Conedera et al. [8]	2006	Alpine Region	Lightning strikes are a significant fire ignition source in specific areas, with distinct characteristics and potential increase due to climate change.
Oliveira et al. [10]	2012	Mediterranean	Random Forest models outperform Multiple Regression for fire occurrence prediction, highlighting precipitation and soil moisture as key factors.
Bilucan et al. [11]	2024	Antalya, Turkey	Elevation seems to be the most critical factor influencing wildfire occurrence, with other topographic and climatic factors like aspect, temperature, precipitation, and wind speed also playing significant roles.

meteorological data, and Machine Learning (ML) algorithms to set up and test pipelines that lead to the prediction and classification of open fires [1], [2].

A. CURRENT UNDERSTANDING OF FIRE DYNAMICS

Despite advances in these areas, significant challenges remain in adapting methodologies to account for different environmental and anthropogenic conditions, particularly in areas less prone to frequent fire situations [3], [4]. To effectively predict these fires, a thorough understanding of their underlying dynamics is essential. The ignition and spread of fires result from a complex interplay of environmental and anthropogenic factors. Key drivers include meteorological conditions, vegetation and fuel characteristics, topography, and human activities. These factors influence both the likelihood of ignition and govern fire intensity and spread potential. Understanding these drivers is crucial for developing accurate predictive models and effective fire management strategies [5], [6], [7].

The relative importance of human and environmental drivers varies by region. Ganteaume et al. [6] emphasized that in the Mediterranean, human-caused fires dominate, often linked to arson or proximity to infrastructure. Environmental factors, such as weather, fuel, and terrain, also significantly influence fire events. Though human activity is a critical factor, natural causes like lightning cannot be overlooked. Conedera et al. [8] examined lightning-induced fires in the Alpine region, revealing that these fires typically occur at higher elevations and steeper slopes. Their research highlighted that climate change, through increased drought frequency, may amplify lightning-induced fires [8], [9].

Given the regional variations in fire drivers, selecting appropriate predictive models is crucial. For instance, Oliveira et al. [10] found that Random Forest models outperformed Multiple Regression in predicting fire occurrences in the Mediterranean, with precipitation and soil moisture as key variables. In regions like Antalya, Turkey, where wildfires are frequent, Bilucan et al. [11] leveraged ensemble machine learning techniques to analyze 13 key fire drivers, identifying elevation as the most critical factor. The complex interaction of these drivers underscores the need for dynamic models capable of adapting to changing conditions. For example, the Mediterranean region's susceptibility to human-caused fires contrasts with areas such as the Alps, where lightning plays a significant role [6], [8].

Furthermore, the diverse methodologies applied in these studies demonstrate the evolving nature of fire prediction techniques. While Oliveira et al. [10] showed the strength of Random Forest models, Bilucan et al. [11] highlighted the effectiveness of ensemble methods visualized by fire susceptibility maps. Clustering techniques, such as k-modes and k-means clustering, have also been employed to analyze spatiotemporal fire data, providing insights into fire patterns while facing challenges in determining optimal cluster numbers and maintaining interpretability [9], [12].

The main literature findings are summarized for better readability in Table 1.

B. ADVANCES AND CHALLENGES IN MACHINE LEARNING FOR FIRE PREDICTION

The application of machine learning in fire prediction has yielded promising results. Models like Random Forest (RF), Support Vector Machines (SVM), and ensemble methods have demonstrated success in integrating meteorological, topographical, and remote sensing data to create fire susceptibility maps [2], [13].

For example, techniques such as Genetic Algorithms (GA) have been employed to optimize feature selection, enhancing model performance by focusing on the most influential factors [13]. Hybrid approaches, such as GA-ANFIS (adaptive neuro-fuzzy inference system), have further improved

TABLE 2. Literature overview of selected studies on machine learning (ML) techniques and fire modelling.

Authors	Year	Area	Method	Main Conclusion
Kalantar et al. [2]	2020	Chaloos Rood watershed, Iran	MARS, BRT, SVM	ML models like MARS, BRT and SVM are promising for forest fire susceptibility prediction
Rodrigues & Riva [20]	2014	Peninsular Spain	ML	ML models can identify human-caused fire risk factors
Tutmez et al. [9]	2017	Antalya, Turkey	K-modes clustering	K-modes clustering is useful for analyzing categorical data in forest fire mapping but may not always find the globally optimal solution.
Hong at al. [13]	2018	Dayu County, China	ML with GA for feature selection	RF outperforms SVM for forest fire susceptibility mapping. GA improves model performance by selecting informative features.
Moayedi et al. [14]	2020	Fire-prone region of Iran	Fuzzy-metaheuristic ensembles (ANFIS with GA, PSO, DE)	Ensemble model with Genetic Algorithm (GA-ANFIS) achieved the best results for forest fire susceptibility mapping.
Gigović et al. [15]	2019	Tara National Park, Serbia	Ensemble Model (SVM-RF)	Ensemble model (SVM-RF) outperforms SVM and RF for forest fire susceptibility mapping.
Mohajane et al. [16]	2021	Northern Morocco	Hybrid Machine Learning (Frequency Ratio with MLP, LR, CART, SVM, RF)	Random Forest-Frequency Ratio (RF-FR) achieved the highest accuracy for forest fire susceptibility mapping.

ML – Machine Learning, MARS – Multivariate Adaptive Regression Splines, SVM – Support Vector Machine, RF – Random Forest, BRT – Boosted Regression Trees, GA – Genetic Algorithms, PSO – Particle Swarm Optimization, DE – Differential Evolution, MLP – Multilayer Perceptron, LR – Logistic Regression, CART – Classification and Regression Tree.

predictive accuracy by combining fuzzy logic systems with ML algorithms [14]

However, these approaches often emphasize prediction accuracy without addressing their scalability to diverse environments or their reliance on extensive high-resolution datasets. Ensemble models like SVM-RF and hybrid frameworks such as RF-FR have shown significant promise, yet their application to regions with minimal natural fire risk and increasing human-induced ignitions remains underexplored [15], [16].

Furthermore, common clustering techniques, such as k-modes clustering, have been utilized to analyze categorical fire data. These methods provide information about fire patterns but face challenges in determining optimal cluster numbers and maintaining interpretability [9]. Such limitations highlight the need for methodologies that can handle the heterogeneity of fire drivers in diverse regions.

Recent works have also explored the use of drone-based AI systems for wildfire monitoring and risk prediction [17], crowd-sourced wildfire spread prediction using smartphones [18], and explainable AI approaches for predicting wildfire spread using convolutional neural networks [19].

To summarize the literature review concerning the use of various ML techniques, we conclude the key features of the research in Table 2.

C. RESEARCH GAPS AND OBJECTIVES

Despite significant advances in GIS and ML techniques for wildfire prediction, several key challenges remain, particularly in regions with low natural wildfire risks, such as the Czech Republic. These challenges include:

1) LIMITED APPLICABILITY OF CURRENT MODELS IN LOW-RISK REGIONS

Many fire prediction models have been developed for fire-prone regions, such as the Mediterranean or California, where fire risks are high due to environmental factors like vegetation, topography, and weather. These models often do not account for regions with lower fire frequencies, where human activity plays a more significant role in ignition. As a result, there is a need to adapt existing models to better handle the unique dynamics of low-risk areas, including human-induced factors such as land-use changes, proximity to urban areas, and socio-economic influences [10], [11].

2) UNDERUTILIZATION OF METEOROLOGICAL AND SOCIO-ECONOMIC DATA

Although meteorological data has been integrated into fire prediction models, many studies still fail to fully leverage the potential of socio-economic and human activity data. In regions like Central Europe, where human influence is a primary fire driver, socio-economic variables such as population density or land-use can significantly impact fire risks. However, these data are often underrepresented in fire prediction models, leading to a gap in understanding the full range of factors contributing to fire occurrences [3], [20].

3) CHALLENGES IN PREDICTION

Many existing fire prediction models rely on observations datasets. The dynamic nature of fire risks, influenced by changing human behavior, urban development, and climate conditions, requires models that can continuously adapt to new information and evolving environmental conditions [2], [13]. Current methods also struggle with incorporating temporal dynamics, such as the seasonality of fire risks and time-of-day factors, which are crucial for understanding human-driven fire events in urban and semi-urban areas.

This study seeks to address these gaps by developing a novel GIS-based and ML-based framework tailored to regions where natural wildfire risks are low, but human activity plays a dominant role. Our specific objectives include:

1) Evaluating the capacity of ML models to predict and characterize fire occurrences driven by diverse environmental and human factors.
2) Investigating the formation of geospatial clusters that reflect distinct fire-causing conditions, enabling targeted fire management strategies.

Our hypothesis is that by analysing spatial and meteorological data from fire risk predictions, additional details like the time of day of fire occurrence can be inferred. This could help us distinguish between modal fires in dry, hot environments and anomalous fires occurring under unfavourable conditions, likely caused mainly by humans. Furthermore, the aim is to develop a procedure both ML-based and GIS-based. The purpose of the procedure is to identify locations with specific fire tendencies.

II. DATA AND METHODS

The overall financial cost of fires in the Czech Republic increased significantly during the period 2001–2020 [21], [22]. Rising temperatures and altered hydrological cycles driven by climate change have increased vegetation flammability [23]. Human factors, such as land-use changes, outdoor activities, and negligence, also contribute to fire risks, while natural causes like lightning remain relevant [24].

The diverse topography and predominantly forested landscape of Czechia create conditions favorable for fire spread. Coniferous forests, dominated by spruce and pine, are widespread, while deciduous forests, such as oak and beech, are more common at lower altitudes [25], [26]. Changes in one aspect of the Czech landscape often influence others, increasing fire risk.

Although the Czech Republic has not yet experienced widespread wildfires, increasing trends and scientific evidence suggest a growing threat [27]. To address this risk, accurate fire prediction models and the exploration of new predictors are essential for effective risk mitigation

A. DATA

This study utilizes fire data for classification, clustering, and explaining fire location. To achieve these goals, the data is categorized into three main groups by role: Fire Data (Target), Predictive Data, and Explanatory Data (Table 3).

All these data sources were merged into a comprehensive dataset on fire occurrences, with meteorological and geographical conditions assigned to each event. For consistency with the predictive models, all datasets were temporally restricted to the period between January 16, 2015, and December 31, 2022.

1) FIRE DATA

Understanding fire occurrence patterns is crucial for producing clear and actionable analytical outputs. This study analyzes data on fires that occurred in the Czech Republic between January 16, 2015, and December 31, 2022, provided by the Fire Rescue Service of the Czech Republic. A focus was placed on the fires in forests, fields, and grasslands, resulting in a sample of 36,780 fire incidents. The dataset includes information on the fire's location, time, and, if determined, the presumed cause of ignition.

TABLE 3. Overview of the data sources and their role in the modelling.

Category	Subsets	Source	Role
Fire occurrence	Hour, Location	Czech Fire Rescue Service	Target
Fire cause/ Trigger	What was the trigger reported by the firefighter dpt.?	Czech Fire Rescue Service	Explanatory
Weather data	Soil temperature, precipitation, field capacity, Haines index	Global Forecast System (GFS) analysis archive	Predictive
Geographical data	Population density, land use type, digital elevation model	Land Survey Office, Czech Statistical Office	Predictive

The causes of fires were identified by the Fire Rescue Service of the Czech Republic according to the Instruction of the Director General [28], which outlines procedures for investigating fire causes. This study specifically examines the time of fire reporting and the reported cause of ignition to investigate factors influencing fire events. Subsequent sections analyzed meteorological and geographical conditions associated with fire ignition.

2) GEOGRAPHICAL DATA

The dataset was enriched with geographic data from official sources, including elevation, land use, and population density.

- **Elevation Data** was retrieved from the 5-meter resolution digital elevation model [29].
- **Land Use Data** was classified as urban, peri-urban (e.g., landfills), or rural (e.g., forests, fields) based on the Fundamental Base of Geographic Data [29].
- **Population Density Data** was obtained from the Czech Statistical Office [30].

3) GLOBAL FORECAST SYSTEM DATA

To identify parameters influencing fire behaviour and patterns, we analysed data from the Global Forecast System (GFS), a weather prediction model developed by the National Oceanic and Atmospheric Administration [31]. We leveraged outputs from the initial model run time, which served as potential predictor variables. The selection of predictors was guided by the literature review and initial exploratory analysis, focusing on the following variables:

- **Soil Temperature (0-6 cm)**: Represents the temperature of the upper soil layer, influencing fire ignition and spread potential, and demonstrating resistance to short-term fluctuations [32].
- **Field Capacity**: Measures the soil's maximum water retention capacity. Drier soils are more prone to ignition, increasing fire risk.

TABLE 4. Overview of the used algorithms and their purpose.

Method	Type	Name	Purpose
ML	Clustering	K-Means	Clustering conditions
		K-Medoids	Clustering conditions
	Classification	Random Forest Gradient Boosting* Ada Boosting Logistic Regression Decision Trees K-Nearest Neighbours Support Vector Machine	Classification of daytime of occurrence
GIS	Clustering	Multivariate	Count – value comparison
		K-medoids	Spatial clustering conditions

* In the more time-consuming cases, Light Gradient Boosting Machine was used.

- **Haines Index (HI, H-index)**: A composite metric of temperature, humidity, and wind speed that assesses fire danger. Higher values indicate drier conditions, stronger winds, and a higher potential for fire spread.
- **Accumulated Precipitation (6, 12, and 24 hours)**: Captures total rainfall within the specified periods preceding a fire event, affecting ignition likelihood.

Meteorological variables were treated individually in exploratory analyses and jointly within ML models to account for complex interactions.

B. METHODS

For clustering and classification, we used a portfolio of methods (Table 4), categorized by their environment and type. Methods were applied either in an integrated development environment (without location data) or within a GIS to identify spatial clusters.

Clustering was performed independently to explore data structures, while classification models were used to predict fire risk based on temporal and environmental factors. The clustering results were not used to refine classification predictions but rather to assess spatial patterns. The multivariate GIS clustering method refers to a spatial analysis technique that evaluates multiple layers (e.g., model accuracy and fire events density) to identify spatial hotspots.

To evaluate the algorithms performance, we used highly interpretable metrics: Accuracy, Precision, Recall, and F1-Score. Their short explanation is in Table 5.

For the clustering tasks, the Silhouette score (equation 1) was used:

$$s(i) = \frac{b(i) - a(i)}{max\,\{b(i) - a(i)\}} \tag{1}$$

where:

TABLE 5. Accuracy metrics used for the classification models performance monitoring.

Name	Equation	
Accuracy	(TN + TP) / TS	What proportion of predictions were correct?
Precision	TP / (TP + FP)	Out of all positive predictions, what proportion were correct?
Recall	TP / (TP + FN)	Out of all actual positive cases, what proportion were correctly identified?
F1-Score	2* (P * R) / (P + R)	How well does the model balance precision and recall?

TP – True Positives, TN – True Negatives, FP – False Positives, FN – False Negatives, TS – Total Samples, P – Precision, and R – Recall.

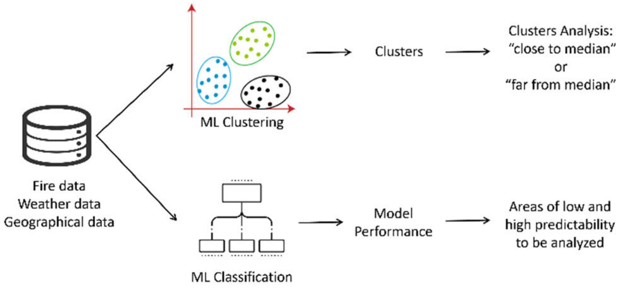

FIGURE 1. Flowchart of the overall workflow.

- *a(i)*: the average distance of the point to all other points in the same cluster. This measures the cohesion of distances within its cluster.
- *b(i)*: The smallest average distance of i to all points in the other clusters. This represents the separation between i and the closest neighbouring cluster.
- The overall Silhouette score for the dataset is the mean of s(i) over all data points.

It is a quantitative measure used to assess the quality of cluster assignments within a dataset [33]. It provides a comparative evaluation of the similarity between a data point and its assigned cluster relative to other clusters. A higher coefficient indicates that a data point is well-matched to its cluster and distinctly separated from neighbouring clusters. Conversely, a lower or negative coefficient suggests potential misclassification or overlapping clusters. The Silhouette Score was computed iteratively to determine the optimal number of clusters, ensuring meaningful spatial groupings.

All statistical calculations were carried out using R software [34], Python libraries, and ArcGIS Pro [35], which was also used for data visualizations and map creation.

To facilitate a comprehensive understanding of the analytical framework employed in this study, the overall methodological workflow is summarized in a schematic flowchart (Fig. 1).

C. DESCRIPTIVE STATISTICS

We employed a linear regression coefficient to overview the trend in fire occurrence at each 0.25-degree grid point of the

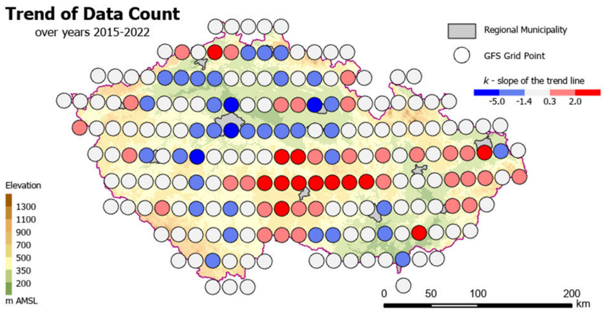

FIGURE 2. Trend in the data count over the years (position of the aggregation points based on the points on the GFS grid) expressed as the slope coefficient *k* of the linear regression.

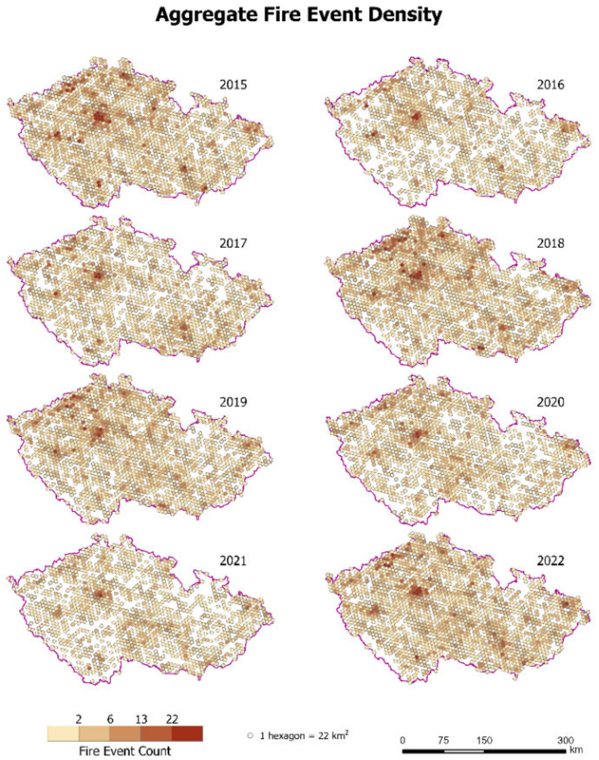

FIGURE 3. Density of aggregate points (fire events) by year using a hexagonal grid.

GFS meteorological model (Fig. 2). This analysis provides a visual representation of regions experiencing increasing or decreasing fire numbers over time.

According to the distribution of fires in the Czech Republic, knowing the population density, the capital city of Prague is noticeable. Then the line of cities in the northwest, where especially in 2015, 2018, 2019 and 2022 the increased number of fires can be assumed (Fig. 3).

1) TIME OF OCCURANCE

The time of day is one of the key parameters. The distribution of the frequency of open fires depending on the calendar month and the time of day is illustrated in Fig. 4. The hour of the day represented a key modelling parameter due to its twofold benefit. Firstly, it is a quantitative variable readily integrated into the model. Secondly, it has the potential to serve as a model output. Predicting the time of day when fires occur helps differentiate human-induced and natural fire triggers, enabling targeted prevention strategies.

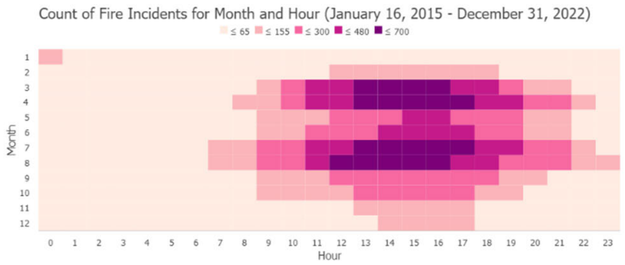

FIGURE 4. Summarized diurnal course of fires for different months.

As evident in Fig. 4, there is a clear connection between the hour of the day, the temperature, and possibly human activity. Therefore, during the modelling application phase, our aim was to predict the time of occurrence. Since it is apparent that there are some fire-prone hours, we opted for using the cyclical time of the day and year to preserve its value. Sine and cosine transformations were chosen over categorical encoding to capture periodicity, allowing models to recognize smooth transitions between hours (e.g., midnight to 1 a.m.) and months (e.g., December to January).

The meteorological data used exhibit a diurnal cycle, suggesting a high correlation with time. Based on the GFS data, we constructed histograms for the analysed meteorological variables (Fig. 5). These histograms indicate that the Haines Index may serve as a relatively strong predictor, providing a solid foundation for assessing conditions conducive to fire occurrence.

Fires with a high Haines Index (HI) tend to occur in the afternoon, aligning with the maximum daily temperatures. This suggests that ideal burning conditions drive these fires, while those with lower HI may arise from non-weather-related factors throughout the day. This knowledge enables us to track the daily progression of fire activity as a function of the HI. Leveraging machine learning methods, we explored downscaling and classifying the Haines Index to use it as an analytical tool to predict various characteristics related to fires.

2) CAUSES AND TRIGGERS

We also examined the cause and description attribute listed in the fire database (Fig. 6). Identifying spatial patterns of specific fire causes, such as lightning or arson, provides essential context for ML model development, ensuring the models can differentiate between natural and human-induced fires.

Since it is clear that of the data where a description or cause is given, the majority is man-made, we include some geographic data in the modelling, such as population density

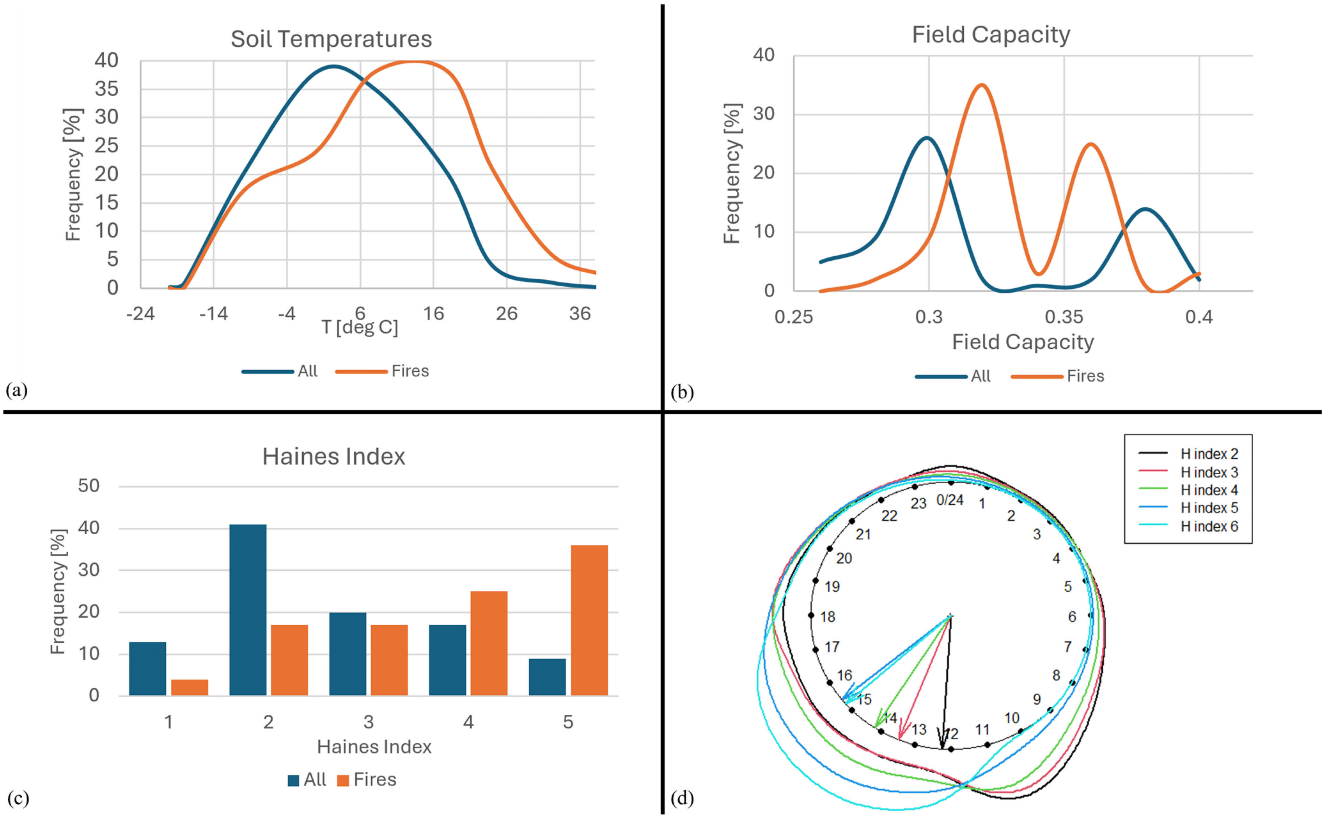

FIGURE 5. **Comparison of the selected predictors related to all meteorological data and fires in the area of interest: (a) Soil temperature, (b) Field capacity, (c) Haines Index, and (d) Haines Index Distribution with the hour of the day showing the most pronounced diurnal cycle of the highest values.**

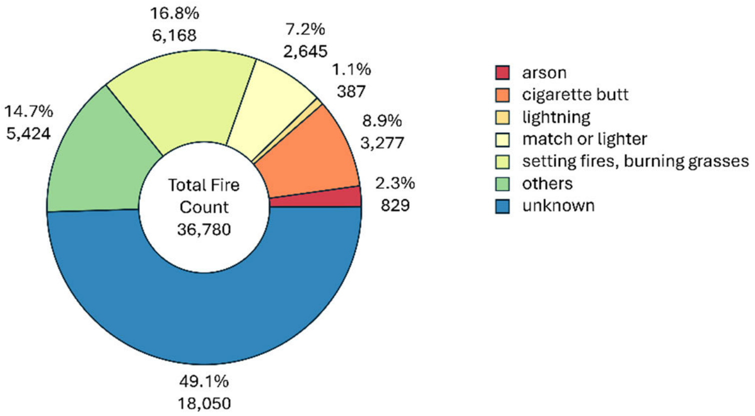

FIGURE 6. **Causes of the fires within the dataset (absolute numbers and their corresponding percentage contributions).**

in a given municipality. However, it is anticipated that in Central Europe in general, it is human activity that is an unpredictable triggering factor for fires. However, this can be linked to time, temperature, or, for example, rainfall, which can significantly determine both fire risk and human activity.

The following figures depict the spatial distribution of fires in the country (Fig. 7 – Fig. 13). The visualization of the map fields was also designed to illustrate the influence of the spatial distribution of fires on geographical characteristics such as elevation and the location of regional municipalities. Fig. 7 highlights the fires attributed to lightning strikes, a purely

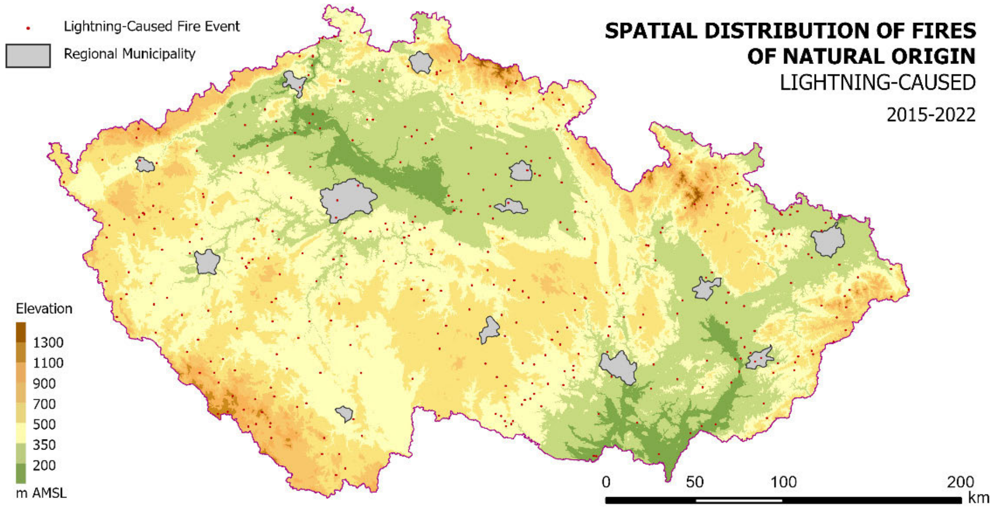

FIGURE 7. Spatial distribution of fires caused by lightning.

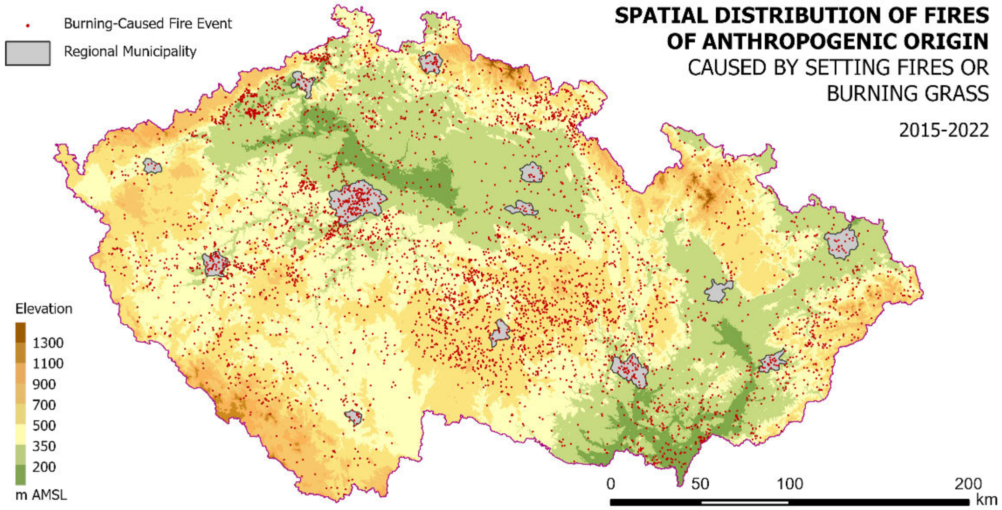

FIGURE 8. Spatial distribution of fires caused by setting fires in nature or in landfills, burning grasses.

natural source. The distribution of these fires appears relatively uniform, with a certain concentration in the southwestern border mountains (Šumava). This low-inhabited region experiences enhanced convection due to the orientation of the mountains and hosts a national park with widespread forested areas.

Human-caused fires, including setting fires in nature or in landfills and burning grasses (Fig. 8) were further depicted. The database contains sub-categories such as burning of branches after forestry work, etc. Fig. 8 shows notable bigger-scale clustering patterns

As analysed in a previous study by Hubáček et al. [36], the central region of the Czech Republic exhibits a distinctive pattern of fires around settlements characterized by smaller towns with rural suburbs, numerous orchards, and gardens. Additionally, there has been significant forest activity in the past due to the drying of forests. This region has experienced pronounced drought conditions. However, as illustrated in Fig. 9 and Fig. 10, smoking and associated cigarettes butts and matches is more an urban issue, and national parks and natural areas are almost bypassed. However, it is one of the most common triggers when identified.

A spatial distribution very similar to that of fires caused by matches or lighters also corresponds to arson-related incidents (Fig. 11). This is a very interesting finding, which may be conditioned by a common behavioural trait of the population in these areas, defining clearly identifiable clusters (e.g. the one in the northwest). Fig. 12 shows the spatial distribution of the other, otherwise classified, human-caused fires, and Fig. 13 shows the spatial distribution of fires caused by unknown causes.

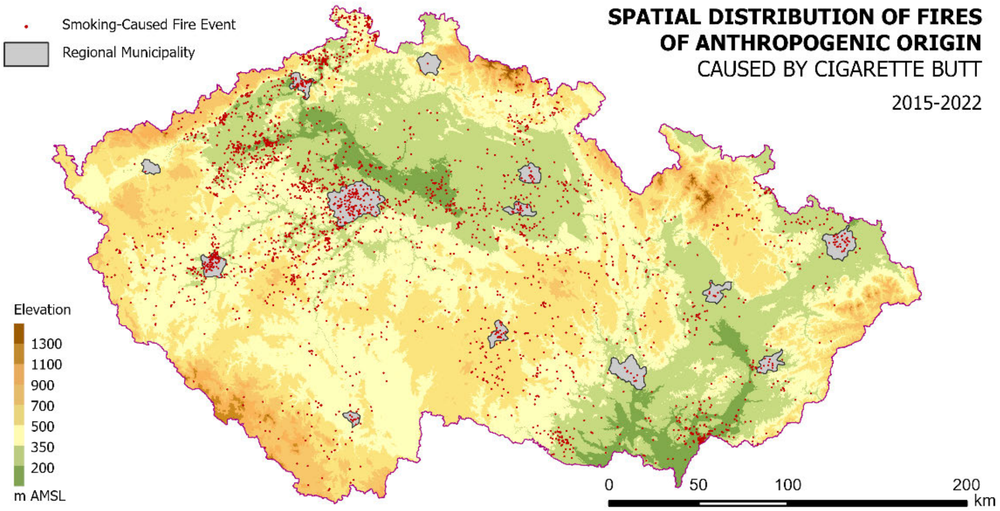

FIGURE 9. Spatial distribution of fires caused by cigarette butt.

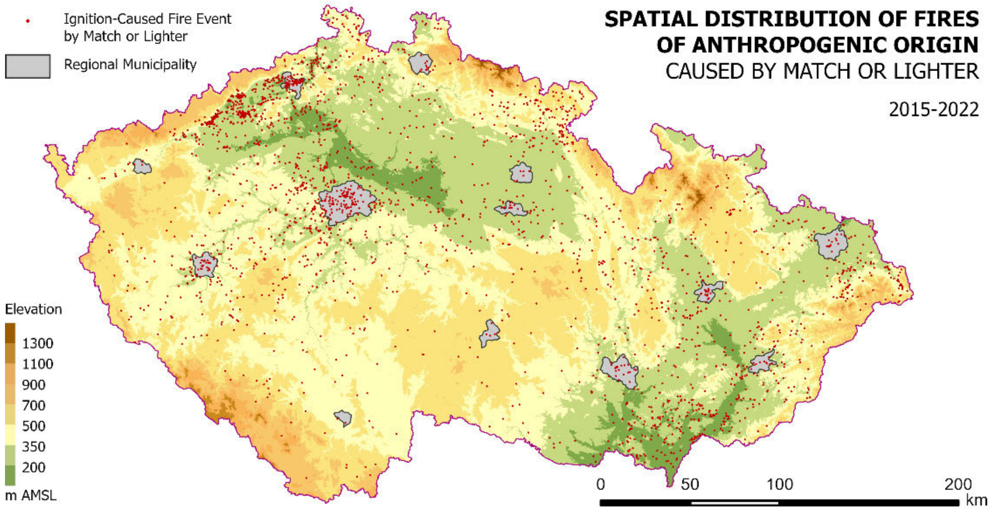

FIGURE 10. Spatial distribution of fires caused by matches or lighters.

As presented in figures 7-13, human activity significantly influences the distribution of fires. Their locations appear to be essentially coincident with the population density. With a knowledge of local geography, it is interesting to note that, for example, in figure 9, the density is higher in the smaller cities of the Northwest. It can be assumed that there is also a significant effect of population composition, standard of living, or lifestyle. The significant number of fires with undetermined causes in the dataset exposed limitations of the data recorded by the fire units. Thus, it was not possible to further classify these events.

III. RESULTS

A. K-MEANS AND K-MEDOIDS CLUSTERING

Our primary goal was to define characteristics for representative (e.g. median-case) fire events and identify those occurring at the boundaries, away from the cluster centroids. To detect outlier fires, we used centroid-based clustering methods like k-means and k-medoids [37]. These methods were preferred over hierarchical clustering because they allowed identification regions where events deviate from the norm by defining clusters with clear centers. While density-based clustering could offer insights into high-density areas, centroid-based methods were more suited to our objectives.

Initial clustering on all fire events resulted in a low Silhouette score (<0.27), indicating that the features did not adequately capture the complexity of fire occurrences. This prompted us to focus on meteorological conditions with continuous variables and filter the input using the categorical Haines Index. The low Silhouette score suggested that our feature set needed refinement to achieve clearer cluster separation. (Fig. 14).

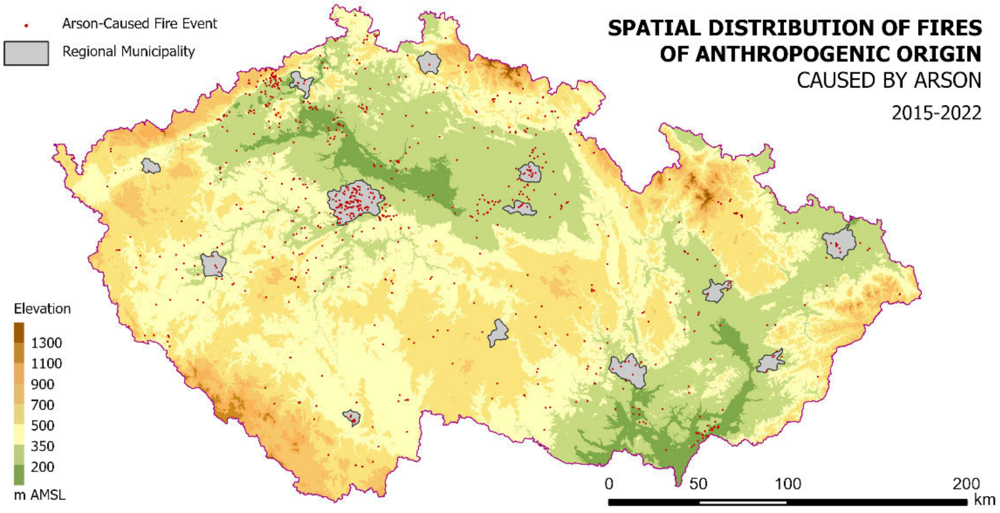

FIGURE 11. Spatial distribution of fires caused by arson.

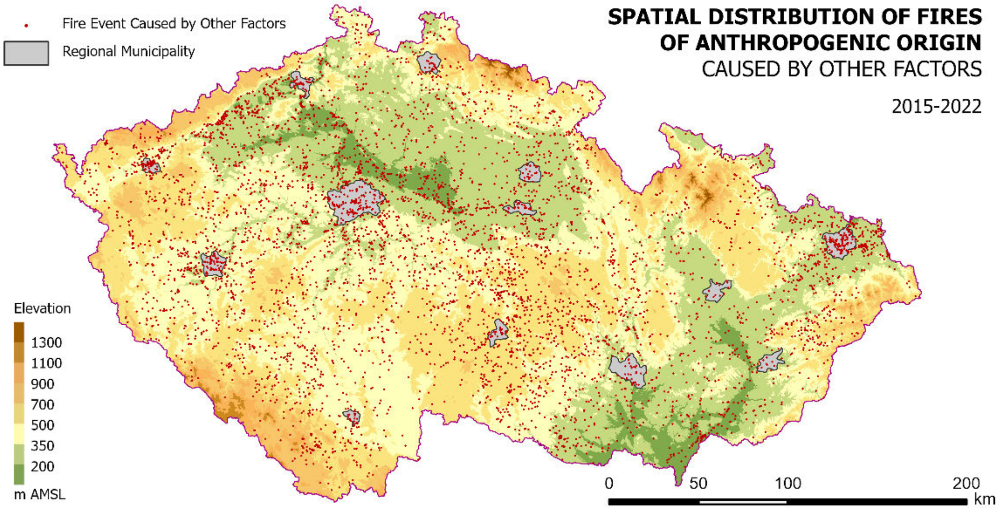

FIGURE 12. Spatial distribution of fires caused by other human actions.

Despite splitting by Haines index, the Silhouette score remained difficult to improve beyond 0.5, suggesting that further adjustments to the feature set or clustering approach might be necessary. As a result, we expanded the dataset by including geographical data, such as population density (pop_density), to assess whether geographical factors could enhance clustering performance. Table 6 summarizes the performance of different clustering iterations.

The analysis summarized in Table 6 explored the clustering of fire events using features like temperature (TMP_L1), population density (pop_density), and temporal aspects (hour/month sine/cosine). The results suggest connections between feature selection, Haines Index, and clustering difficulty. Although the Silhouette Score suggested some clustering challenges, the selected features provided useful information about the suitability of selected features.

The last cluster's Silhouette Score significantly decreased with Haines Index = 6, indicating complexity in the clustering process, but also suggesting inappropriate features for high-risk periods. The anticipation is that the sine function effectively separates the summer months (high Haines Index values) from the rest of the year, leading to better clustering with Haines Index equal to 6. The cosine function, with its focus on spring/fall differences, may not capture this seasonal pattern as effectively.

The impact of Haines Index on clustering is noteworthy. In some cases, higher Haines Index values correspond to better Silhouette Scores. This suggests that categorizing fire events based on a fire risk index could be beneficial. However, it is crucial to avoid overfitting the data by using too many clusters.

To improve clustering performance, future studies could explore the following.

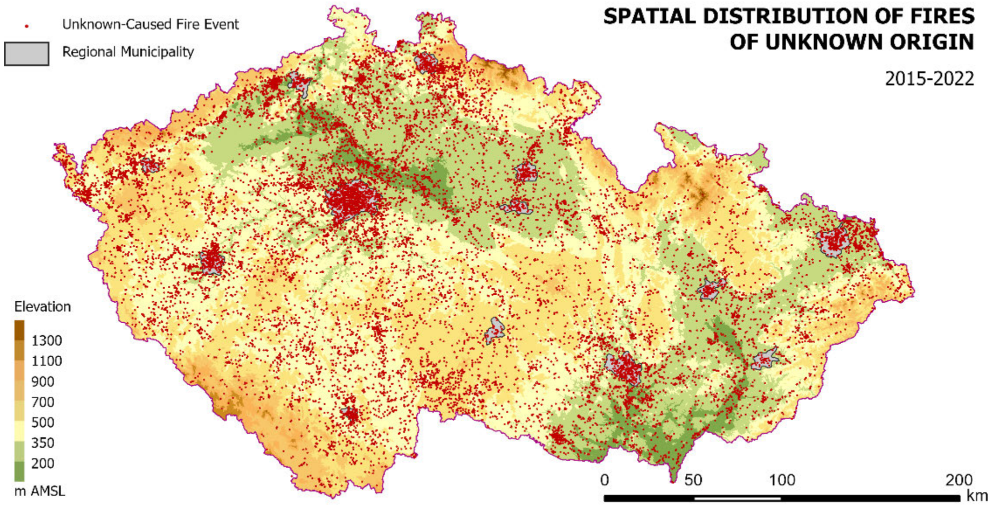

FIGURE 13. Spatial distribution of fires caused by unknown cases.

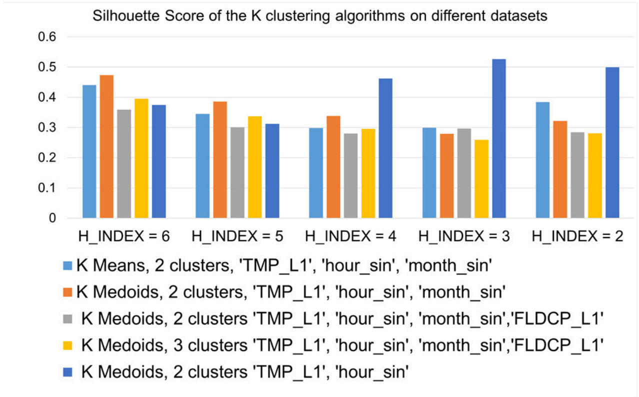

FIGURE 14. Tested configurations of clustering algorithms, including type (Means/Medoids), number of clusters, and predictor sets. 'TMP_L1' represents soil temperature, 'hour/month_sin' denotes the sine transformation of time, and 'FLDCP_L1' refers to field capacity as a parameter of first-layer soil wetness.

TABLE 6. Clustering iterations performance metrics applied on the different datasets.

Feature Set	Haines Index	Silhouette Score
TMP_L1, pop_density, hour_sin	2	0.45
	3	0.48
	4	0.47
	5	0.40
	6	0.55
TMP_L1, pop_density, hour_sin, month_sin	2	0.31
	3	0.35
	4	0.40
	5	0.44
	6	0.50
TMP_L1, pop_density, hour_sin, month_cos	2	0.37
	3	0.35
	4	0.35
	5	0.41
	6	0.19

TMP_L1 – soil temperature, pop_density – population density, hour/month_sin/cos – sine/cosine function of the time/month.

- Inclusion of additional features: Investigating features beyond the ones used here, such as vegetation type or topography [5], might provide a richer but also more complex dataset for clustering.
- Dimensionality reduction techniques: Techniques like Principal Component Analysis (PCA) could help identify a smaller set of features that effectively represent the data for clustering purposes.
- Analysis of high-scoring clusters: Analysing the characteristics of clusters with higher Silhouette Scores can provide explanatory case studies, the specific fire event conditions that are easier to distinguish based on the chosen features.

Finally, the most successful algorithm was applied and viewed the content of the triggers by clusters (Table 7). The table suggests partially correct assumption that higher Haines Index might overlap with specific causes. It shows a trend in which the Silhouette Score indicating better cluster separation increases with Haines Index. This suggests that for higher fire risk conditions (higher Haines Index), the clustering algorithm could perform better at grouping fire events according to their causes.

Specifically, for Haines Index = 6 (highest risk), the most dominant causes could indicate that during periods of high fire danger, human activity such as smoking ("Cigarette butt" and "Match or lighter") or accidental setting fire and burning (e.g., during gardening and forestry works) becomes a more prominent factor. A similar trend can also be observed in other causes, which are not dominant, but in their sum, an increasing number of cases with higher Haines Index can be observed. Additionally, the dry conditions associated with

TABLE 7. Silhouette score of the K-Medoids clustering algorithms based on the Haines index (HI) value with the most frequent causes.

HI	Silhouette score	Number of fires by cause							
		Total	Arson	Cigarette butt	Lightning	Match or lighter	Setting fire, burning grass	Others	Unknown
2	0.45	1785	71	131	19	115	274	324	851
3	0.47	6279	170	438	98	422	987	1104	3060
4	0.47	6366	157	506	73	515	964	1063	3088
5	0.40	9165	207	791	109	701	1395	1329	4633
6	0.55	13185	224	1411	88	892	2548	1604	6418

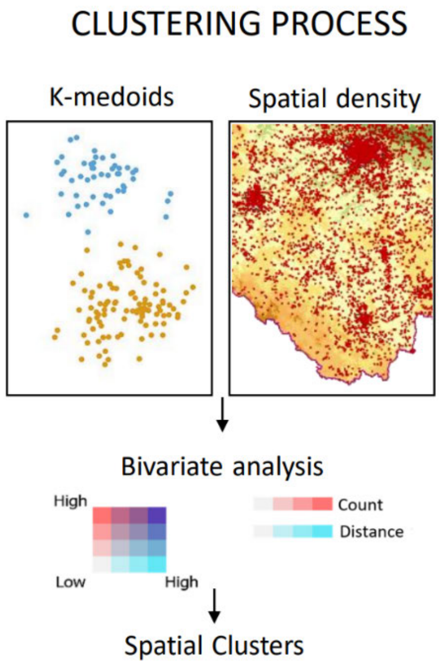

FIGURE 15. Tested configurations of clustering algorithms, including type (Means/Medoids), number of clusters, and predictor sets. 'TMP_L1' represents soil temperature, 'hour/month_sin' denotes the sine transformation of time, and 'FLDCP_L1' refers to field capacity as a parameter of first-layer soil wetness.

high HI might make it harder to pinpoint the exact cause, leading to more "Unknown" classifications.

"Arson" and "Lightning" appear as causes across all Haines Index values, suggesting that these factors could be less influenced by overall fire risk conditions. Arson might be a more consistent human-caused factor, while lightning strikes can occur regardless of fire danger.

The high number of "Unknown" causes across all Haines Index values poses a limitation of the data. This could cover significant groups that would skew the statistics of the causes towards different ratios.

Despite weakened clustering due to human influence, our method (Fig. 15) proved its ability to spatially separate fire causes (not included in clustering). Using different data samples and representative fires, fire risk indices, etc., different clusters and homogeneous regions would separate.

FIGURE 16. Military training area Boletice identified by the high distance from the centroids during the clustering and bivariate analysis. (Basemap source: [29])

TABLE 8. Number of all cases in highlighted cluster in boletice area and number of extraordinary (military) causes.

HI	Number of Cases	Explosives or Military Causes
2	3	3
3	17	13
4	29	15
5	37	13
6	5	4

The proposed method identified, among others, a cluster in the South Bohemian region (Boletice area, Fig. 16) with high fire count and large distances from the centroid (purple color of the bivariate cluster). This suggests atypical, localized fires.

Analysis revealed a high proportion of single-cause fires in this area, with all clusters having above-average distances from the centroid, indicating rather anomalous parameter combinations. Interestingly, there is a military training area. Inspired by this cluster, we investigated the number of fires (Table 8) caused by military activities (bullets, explosives,

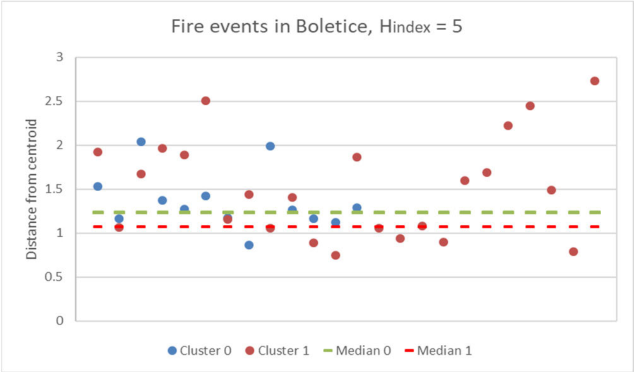

FIGURE 17. Distances from the centroid from boletice area for K-medoids cluster, H-index 5 indicating higher fire risk.

sparks, etc.) the causes were manually retrieved from the description of the event.

In all cases, the median distance from the centroid was higher than the median in the Czech Republic. As an example, we present the cluster for H-index = 5 (Fig. 17), i.e. higher fire risk.

All median distances from the centroids were higher than the national median for Czechia, indicating that fires in this location are less representative based on the predictors. These findings suggest that our proposed method can effectively highlight spatial clusters of fires with common causes, as evidenced by a high proportion attributed to military activities. This is particularly true under the least favourable conditions for fire occurrence. However, under more favourable conditions, fires may be ignited by a variety of causes.

B. CLASSIFICATION METHODS

Besides clustering, we employed supervised ML classification methods to predict the time of day with the highest probability of fire occurrence. We chose this target variable due to:

- Quantitativeness: The time of occurrence of fire can be easily measured and represented numerically.
- Predictable Distribution: The distribution of fire occurrences throughout the day likely exhibits patterns that can be learned by ML models.
- Potential Applications: Predicting peak fire occurrence times can be valuable for resource allocation and fire prevention strategies. It can also reveal unusual events with different timely patterns.

The day was segmented into four equal time periods (0-6, 6-12, 12-18, 18-24 UTC) corresponding to the GFS run steps. We employed stratified 80/20 train-test data splitting to ensure a balanced representation of fire occurrences across all time periods in the training data. Hyperparameter tuning for each algorithm was performed using Randomized Search with a 10-fold cross-validation. The following ML classification algorithms were evaluated (Fig 18): Random Forest, Gradient Boosting, AdaBoost 2, Logistic Regression, Decision Trees, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM).

While all models in Fig. 16 exhibited limitations in predicting the time interval of a forest fire, analysing their performance through precision, recall, and F1 scores reveals some hints to explore further. Gradient Boosting stands out with the highest accuracy and precision, indicating a strong ability to identify true positives. However, its lower F1 score suggests that it might be missing some actual fires (false negatives) due to its conservative approach. KNN's success likely stems from its ability to handle complex data by leveraging similarities in past fire events. Conversely, SVM's performance suffered because it struggled with data with unclear boundaries, which might be the case for fire occurrence times. These observations could guide further considerations about the nature of the prediction task when selecting a machine learning model.

Although the initial results were not suitable for building a complex hierarchical model, our aim was to improve model performance by incorporating additional geographic information about population density and investigating the impact of data splitting. To assess the influence of data partitioning on model robustness, we employed a repeated 100-fold cross-validation approach for the Random Forest model. This technique involves splitting the data into training on 100 random subsets (80/20 ratio) and monitoring the accuracy to estimate the influence of the feature distribution within the training and testing set.

Random Forest, which leverages an ensemble of decision trees and utilizes a "majority vote" approach for classification, is typically less susceptible to outliers compared to some other models. This characteristic makes it a suitable choice to explore the impact of data partitioning in this context. For a dataset that contains soil temperature, field capacity, population density, and accumulated precipitation, the Random Forest model achieved performance metrics within the range presented in Table 9.

Table 9 highlights the variability in performance metrics even when using only training fires. This suggests a potential presence of two fire sub-groups within the data:

- Representative Subset: This subset might exhibit consistent fire behaviour captured by the model, leading to some stability in performance metrics across training runs.
- Heterogeneous Subset: This subset might consist of more diverse fire events, leading to greater variation in model performance across different training data splits.

We hypothesize that the representative subset might be characterized by the following.

- Human-caused fires.
- Occurrence during appropriate weather conditions (higher temperatures, low humidity).
- Occurrence during periods of high human activity (potentially near settlements).

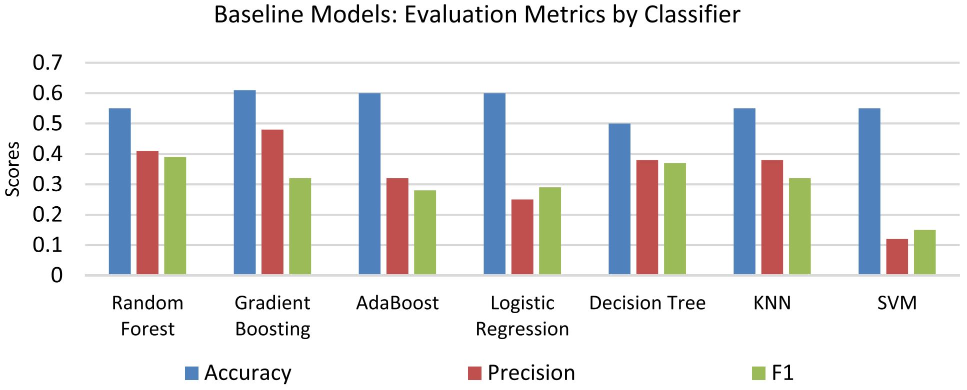

FIGURE 18. Performance metrics of the baseline models.

TABLE 9. Performance metrics for the 100 random train/test splits.

	Accuracy	Precision	Recall	F1 Score	AUC
Range	0.53-0.63	0.51-0.61	0.52-0.60	0.52-0.58	0.69-0.76

AUC – Area Under the Curve.

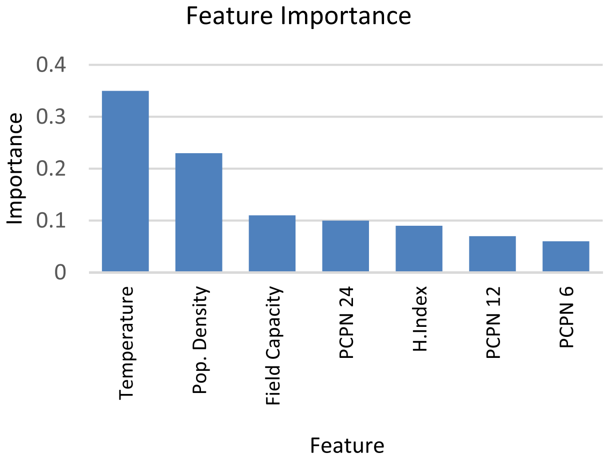

FIGURE 19. Importance of features (predictors) for the Random Forest classification model that predicts the time of the fire occurrence. PCPN 24/12/6 stands for total precipitation over a 24/12/6-hour period.

However, the feature importance analysis provides a more consistent pattern across model training runs (Fig. 19). Random Forest consistently identifies temperature and population density as the most important features, followed by field capacity, precipitation, and the Haines Index.

We can evaluate that precipitation (all three values) holds an important value together with other predictors. To address this, future studies could explore techniques like Principal Component Analysis (PCA) to identify features with reduced correlation or create aggregate measures for features like precipitation. However, this could reduce interpretability of results.

The following figures (Fig. 20) present the accuracy of each method on a 0.25-degree grid depending on fire density (related to the unit area of 0.25 degrees square). The ML algorithms were trained on data from all other grid points within the country and tested on the specific point of interest. This leave-one-out approach enabled us to identify three primary scenarios:

1) High-density, high-accuracy fire locations: Regions with frequent fire occurrences often exhibited high prediction accuracy, suggesting strong correlations with broader spatial patterns.
2) High-density, low-accuracy fire locations: Despite the high fire frequency, certain areas showed lower prediction accuracy, indicating the potential influence of complex local factors.
3) Low-density fire locations: Due to limited data points, the predictive performance in these regions is less reliable and requires case-specific analysis.

Specifically for the Random Forest (RF) and Gradient Boosting (GB) models, they performed quite similarly to analyzing the influence of environmental and anthropogenic factors on fire risk. Visualizations of the ML outputs revealed a connection between the number of fires in a location and the accuracy of the models. Locations with a high frequency of fires often exhibited higher model accuracy (RF/GB). This suggests that these areas likely have more consistent fire drivers related to the factors considered (e.g., consistently high temperatures, low precipitation). Consistent patterns allow models like RF and GB to learn effectively and achieve high accuracy in predicting fire events.

Conversely, locations with frequent fires but low model accuracy (RF/GB) could indicate a diversity of fire causes not well captured by the chosen environmental and anthropogenic factors. Alternatively, limitations in data quality or the exclusion of spatial relationships between fires (presumably nuances in human activity) could contribute to lower accuracy.

Finally, locations with low fire counts presented a challenge for the models due to limited data. With fewer fire events, models struggle to learn robust patterns, and random chances can play a larger role in predictions.

These findings confirm common assumptions of human-caused fires but also emphasize the importance of

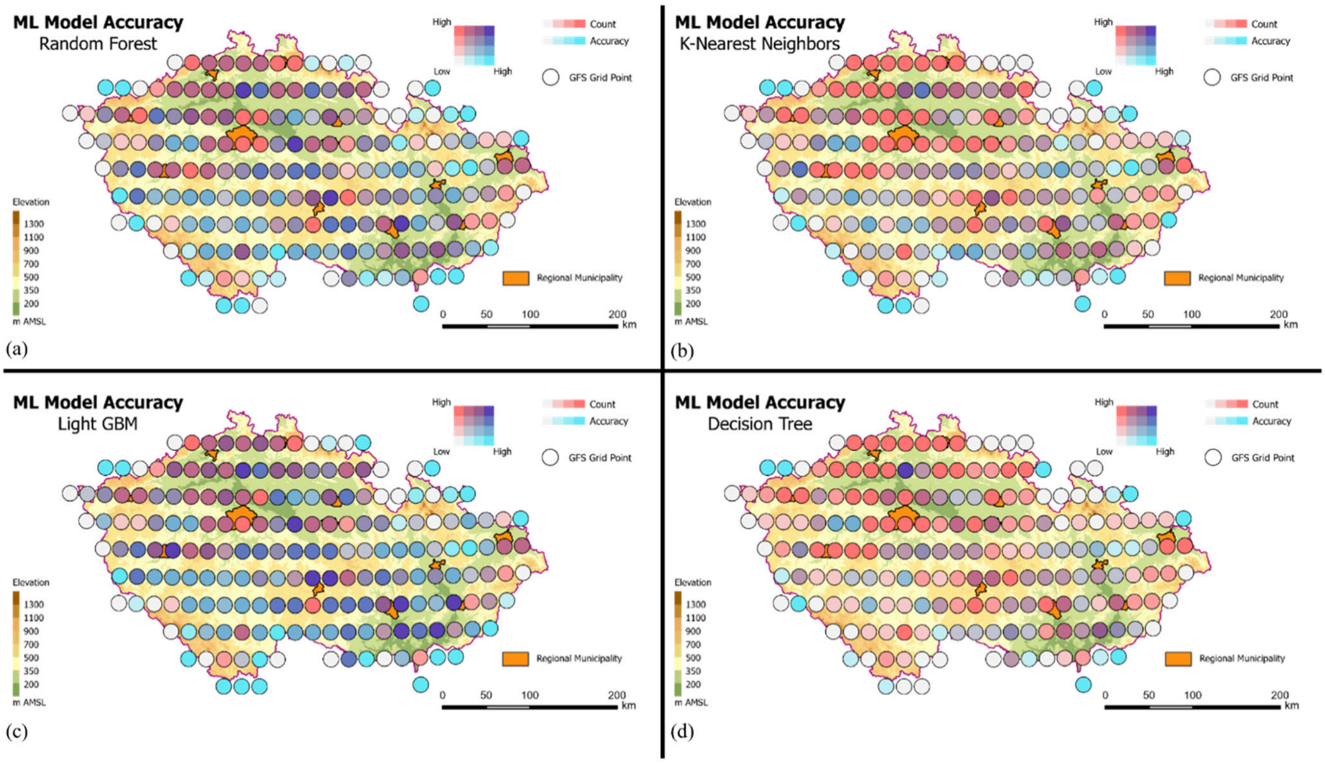

FIGURE 20. **Accuracy of the Random Forest (a); K-Nearest Neighbors (b); Light Gradient Boosting (c) and Decision Tree (d) for each point of 0.25 degrees grid, predicting time of fire occurrence, with accuracy depending on the fire count (density) within each grid cell.**

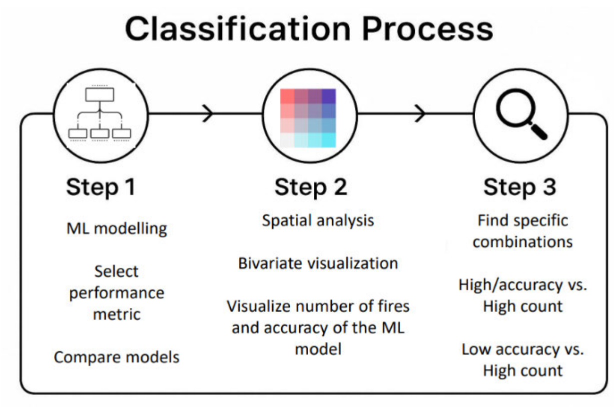

FIGURE 21. **Procedure of the ML diagnostics of the area based on historical data.**

acknowledging limitations. The low accuracy of the model in some locations suggests the influence of factors beyond those quantified in this study. Additionally, it is crucial to interpret model results alongside expert knowledge about fire causes in areas with infrequent fires.

Based on the classification performed and inherent analysis, we propose the procedure to exploit ML classification for determination of the local complexity (Fig. 21).

We suggest continued exploration of machine learning (ML) methods for fire risk location definition. When combined with advanced data, such as output from local meteorological models, these methods could effectively classify fire features under well-understood conditions. These are very often rural areas with well-developed infrastructure, or local settlements directly bordering with natural areas. However, in cases where ML models perform poorly, it implies that even human understanding of fire causes in that location may be limited by the available data. Especially poor performance can be generally seen in proximity of big cities with high fluctuation of the inhabitants. Due to the dataset size, varying accuracy can also be spotted in mountainous and bordering areas. We emphasize that the success of such data-driven ML approaches hinges on both high-quality data and a robust methodological framework..

IV. DISCUSSION

This study evaluated the potential of sequential machine learning and GIS techniques for identification of fire occurrence patterns in the Czech Republic by developing a classification and clustering framework. Unlike previous studies that focus predominantly on arid, subtropical, or Mediterranean regions, this research introduces a novel perspective within the central European (low-risk region) context. Key contributions include:

- Applying ML techniques to a less-studied geographical area (Czechia) with distinct fire dynamics.
- Integrating and sequentially utilizing ML modelling with GIS-based bivariate visualization to enhance interpretability.
- Categorizing input data using the Haines Index and analysing outputs based on reported fire causes to uncover patterns and drivers of fire risk.

Using fire occurrence data and meteorological predictions, the study aimed to identify factors influencing fire events and predict periods of heightened fire risk, such as the time of day with the highest likelihood of fire occurrence.

A. KEY FINDINGS

- **Clustering**: K-Means and K-Medoids established feature-fire event connections, but clear separation was challenging, leading to low Silhouette scores. This suggests feature limitations rooted in the inherent complexity of the predictor-target relationship, which results in significant group overlap and requires further refinement before operational deployment. Better clustering may require additional features (e.g., vegetation type) and dimensionality reduction (e.g., PCA). Notably, higher Haines Index values corresponded to better cluster separation, validating the use of risk indices for grouping similar events. The distance from the centroid effectively identified homogeneous areas, which allowed detection of military areas with unusual patterns (explosives/hot bullets) when combined with spatial clustering. In our case study, K-Medoids combined with spatial clustering detected military areas with unusual fire patterns, caused by explosives or hot bullets.
- **Classification**: Supervised ML methods yielded moderate performance in predicting the time of day with the highest fire risk. Variability in performance suggests potential sub-groups within the data: a representative subset with consistent behaviour and a heterogeneous subset with more diverse events. Nevertheless, temperature and population density emerged as the most relevant factors for predicting the time of fire occurrence.

B. PRACTICAL IMPLICATION

- **Clustering-based approach**: This approach, once fully developed, can become an actionable tool for assessing fire risks. By grouping fire events based on similar characteristics identified through clustering (e.g. high Haines Index and specific causes), forest management authorities can prioritize areas for targeted fire prevention efforts. In the final stage, it can serve as a tool for retrospective estimation of the likely fire causes.
- **Classification-based approach**: This approach has the potential to project a fire risk model. By incorporating real-time weather data and population density information into the model, authorities can generate real-time fire risk maps. These maps can be used to alert fire units in high-risk areas during peak fire times (late afternoons with high temperatures). It might also serve as a diagnostic tool for area complexity. The performance of the ML model itself can be considered as a feature of the area. Highly predictable fire events, likely due to natural causes and influenced primarily by weather and geographical data, would result in simpler statistical characteristics within the model. Conversely, complex statistical characteristics within the model would indicate a more challenging area to model, potentially due to the higher influence of diverse factors like human activity.

V. CONCLUSION AND FUTURE WORK

The current lack of comprehensive data on human activity and fine-scale weather patterns limited the study. Future work should explore the incorporation of these factors to improve model performance. As for previous studies, we suggest focusing on:

*1)*We suggest identifying and integrating data on vegetation types, conditions, and susceptibility to fire risk based on historical fire patterns and the classifications of forest and agricultural experts.

*2)*We suggest developing a more granular classification system for human activity, especially focusing on the work schedules of gardening and forest workers in the region as a potential major fire risk factor.

*3)*We recommend utilizing high-resolution weather data to capture local weather patterns and fluctuations to improve the accuracy of the model. This is crucial for operational ML models used in real-time fire risk assessment.

*4)*The high number of ''unproven'' and ''unidentified'' events points to the need for more precise data on the causes of fires to obtain their full explanatory potential.

*5)*Exploring more advanced models or hierarchical approaches could potentially capture the complex interactions between various factors that influence fire occurrences.

*6)*Integrating spatial data and techniques can help incorporate the influence of human activity patterns on fire risk. However, we suggest incorporating data on human activity in advance so that these relationships can be captured right during ML modelling and exploratory analysis.

In general, this study demonstrates the potential of ML for analysing fire occurrence patterns and classifying the homogeneity of fire risk. By addressing data limitations and exploring more advanced techniques, future research can lead to the development of even more robust models for fire classification, clustering, prediction, but the most likely we anticipate further categorization of fires and demographics to target precautions and policies.

AUTHOR'S CONTRIBUTIONS

Conceptualization: David Sládek and Filip Dohnal; Data curation: František Paulus, Martin Molek, and Tomáš Zeman; Formal analysis: David Sládek, Filip Dohnal, Jiří Neubauer, and Tomáš Zeman; Funding acquisition: David Sládek, František Paulus, and Tomáš Zeman; Methodology: David Sládek and Filip Dohnal; Supervision: Tomáš Zeman; Visualization: David Sládek, Filip Dohnal, and Jiří Neubauer; Writing–original draft: David Sládek and Filip Dohnal; Writing–review and editing: František Paulus, Jiří Neubauer, Martin Molek, and Tomáš Zeman. All authors have read and agreed to the published version of the manuscript.

REFERENCES

[1] F. Huot, R. L. Hu, N. Goyal, T. Sankar, M. Ihme, and Y.-F. Chen, ‘‘Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data,’’ *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4412513, doi: 10.1109/TGRS.2022.3192974.

[2] B. Kalantar, N. Ueda, M. O. Idrees, S. Janizadeh, K. Ahmadi, and F. Shabani, ‘‘Forest fire susceptibility prediction based on machine learning models with resampling algorithms on remote sensing data,’’ *Remote Sens.*, vol. 12, no. 22, p. 3682, Nov. 2020, doi: 10.3390/rs12223682.

[3] C. Mambile, S. Kaijage, and J. Leo, ‘‘Application of deep learning in forest fire prediction: A systematic review,’’ *IEEE Access*, vol. 12, pp. 190554–190581, 2024, doi: 10.1109/ACCESS.2024.3515215.

[4] J. W. Muhs, M. Parvania, and M. Shahidehpour, ‘‘Wildfire risk mitigation: A paradigm shift in power systems planning and operation,’’ *IEEE Open Access J. Power Energy*, vol. 7, pp. 366–375, 2020, doi: 10.1109/OAJPE.2020.3030023.

[5] E. Chuvieco, I. Aguado, S. Jurdao, M. L. Pettinari, M. Yebra, J. Salas, S. Hantson, J. de la Riva, P. Ibarra, M. Rodrigues, M. Echeverría, D. Azqueta, M. V. Román, A. Bastarrika, S. Martínez, C. Recondo, E. Zapico, and F. J. Martínez-Vega, ‘‘Integrating geospatial information into fire risk assessment,’’ *Int. J. Wildland Fire*, vol. 23, no. 5, pp. 606–619, Aug. 2014, doi: 10.1071/wf12052.

[6] A. Ganteaume, A. Camia, M. Jappiot, J. San-Miguel-Ayanz, M. Long-Fournel, and C. Lampin, ‘‘A review of the main driving factors of forest fire ignition over Europe,’’ *Environ. Manage.*, vol. 51, no. 3, pp. 651–662, Mar. 2013, doi: 10.1007/s00267-012-9961-z.

[7] C. A. T. Cortes, S. Thurow, A. Ong, J. J. Sharples, T. Bednarz, G. Stevens, and D. D. Favero, ‘‘Analysis of wildfire visualization systems for research and training: Are they up for the challenge of the current state of wildfires?’’ *IEEE Trans. Vis. Comput. Graph.*, vol. 30, no. 7, pp. 4285–4303, Jul. 2024, doi: 10.1109/TVCG.2023.3258440.

[8] M. Conedera, G. Cesti, G. B. Pezzatti, T. Zumbrunnen, and F. Spinedi, ‘‘Lightning-induced fires in the Alpine region: An increasing problem,’’ *Forest Ecol. Manage.*, vol. 234, p. S68, Nov. 2006, doi: 10.1016/j.foreco.2006.08.096.

[9] B. Tutmez, M. G. Ozdogan, and A. Boran, ‘‘Mapping forest fires by nonparametric clustering analysis,’’ *J. Forestry Res.*, vol. 29, no. 1, pp. 177–185, Jan. 2018, doi: 10.1007/s11676-017-0417-4.

[10] S. Oliveira, F. Oehler, J. San-Miguel-Ayanz, A. Camia, and J. M. C. Pereira, ‘‘Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random forest,’’ *Forest Ecol. Manage.*, vol. 275, pp. 117–129, Jul. 2012, doi: 10.1016/j.foreco.2012.03.003.

[11] F. Bilucan, A. Teke, and T. Kavzoglu, ‘‘Susceptibility mapping of wildfires using XGBoost, random forest and AdaBoost: A case study of Mediterranean ecosystem,’’ in *Recent Research on Geotechnical Engineering, Remote Sensing, Geophysics and Earthquake Seismology* (Advances in Science, Technology & Innovation). Cham, Switzerland: Springer, 2024, pp. 99–101, doi: 10.1007/978-3-031-48715-6_22.

[12] L. Wu and A. Ren, ‘‘Urban fire risk clustering method based on fire statistics,’’ *Tsinghua Sci. Technol.*, vol. 13, no. 1, pp. 418–422, Oct. 2008, doi: 10.1016/S1007-0214(08)70184-6.

[13] H. Hong, P. Tsangaratos, I. Ilia, J. Liu, A.-X. Zhu, and C. Xu, ‘‘Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. The case of dayu county, China,’’ *Sci. Total Environ.*, vol. 630, pp. 1044–1056, Jul. 2018, doi: 10.1016/j.scitotenv.2018.02.278.

[14] H. Moayedi, M. Mehrabi, D. T. Bui, B. Pradhan, and L. K. Foong, ‘‘Fuzzy-Metaheuristic ensembles for spatial assessment of forest fire susceptibility,’’ *J. Environ. Manage.*, vol. 260, Apr. 2020, Art. no. 109867, doi: 10.1016/j.jenvman.2019.109867.

[15] L. Gigović, H. R. Pourghasemi, S. Drobnjak, and S. Bai, ‘‘Testing a new ensemble model based on SVM and random forest in forest fire susceptibility assessment and its mapping in Serbia’s tara national park,’’ *Forests*, vol. 10, no. 5, p. 408, May 2019, doi: 10.3390/f10050408.

[16] M. Mohajane, R. Costache, F. Karimi, Q. Bao Pham, A. Essahlaoui, H. Nguyen, G. Laneve, and F. Oudija, ‘‘Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area,’’ *Ecol. Indicators*, vol. 129, Oct. 2021, Art. no. 107869, doi: 10.1016/j.ecolind.2021.107869.

[17] C. A. S. Lelis, J. J. Roncal, L. Silveira, R. D. G. De Aquino, C. A. C. Marcondes, J. Marques, D. S. Loubach, F. A. N. Verri, V. V. Curtis, and D. G. De Souza, ‘‘Drone-based AI system for wildfire monitoring and risk prediction,’’ *IEEE Access*, vol. 12, pp. 139865–139882, 2024, doi: 10.1109/ACCESS.2024.3462436.

[18] N. Bogdos and E. S. Manolakos, ‘‘Crowd-sourced wildfire spread prediction with remote georeferencing using smartphones,’’ *IEEE Access*, vol. 7, pp. 102102–102112, 2019, doi: 10.1109/ACCESS.2019.2931456.

[19] M. Marjani, M. Mahdianpari, S. A. Ahmadi, E. Hemmati, F. Mohammadimanesh, and M. S. Mesgari, ‘‘Application of explainable artificial intelligence in predicting wildfire spread: An ASPP-enabled CNN approach,’’ *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024, doi: 10.1109/LGRS.2024.3417624.

[20] M. Rodrigues and J. de la Riva, ‘‘An insight into machine-learning algorithms to model human-caused wildfire occurrence,’’ *Environ. Model. Softw.*, vol. 57, pp. 192–201, Jul. 2014, doi: 10.1016/j.envsoft.2014.03.003.

[21] F. Jurečka, M. Možný, J. Balek, Z. Žalud, and M. Trnka, ‘‘Comparison of methods for the assessment of fire danger in the Czech republic,’’ *Acta Universitatis Agriculturae Et Silviculturae Mendelianae Brunensis*, vol. 67, no. 5, pp. 1285–1295, Oct. 2019, doi: 10.11118/actaun201967051285.

[22] H. Nedělníková, ‘‘Statistical yearbook 2001–2020. Ministry of the interior,’’ Directorate General of the Fire Rescue Service of the Czech Republic, Statistical Yearbook 2001-2020, Prague, Czech Republic, 2021.

[23] M. Trnka et al., ‘‘Observed and expected changes in wildfire-conducive weather and fire events in peri-urban zones and key nature reserves of the Czech Republic,’’ *Climate Res.*, vol. 82, pp. 33–54, Nov. 2020. [Online]. Available: https://www.int-res.com/abstracts/cr/v82/p33-54/

[24] N. Fernandez-Anez et al., ‘‘Current wildland fire patterns and challenges in Europe: A synthesis of national perspectives,’’ *Air, Soil Water Res.*, vol. 14, pp. 1–19, 2021, doi: 10.1177/11786221211028185.

[25] T. Pánek and V. Kapustová, ‘‘Long-term geomorphological history of the Czech Republic,’’ in *Landscapes and Landforms of the Czech Republic* (World Geomorphological Landscapes). Cham, Switzerland: Springer, 2016, pp. 29–39, doi: 10.1007/978-3-319-27537-6_4.

[26] M. Adámek, Z. Jankovská, V. Hadincová, E. Kula, and J. Wild, ‘‘Drivers of forest fire occurrence in the cultural landscape of central Europe,’’ *Landscape Ecol.*, vol. 33, no. 11, pp. 2031–2045, Nov. 2018, doi: 10.1007/s10980-018-0712-2.

[27] M. Mozny, M. Trnka, and R. Brázdil, ‘‘Climate change driven changes of vegetation fires in the Czech republic,’’ *Theor. Appl. Climatol.*, vol. 143, nos. 1–2, pp. 691–699, Jan. 2021, doi: 10.1007/s00704-020-03443-6.

[28] *The Collection of Internal Management Acts of the Director General of the Fire Rescue Service of the Czech Republic: Procedure of the Czech Republic Fire Rescue Service to Determine the Causes of Fires*, Fire Rescue Service of the Czech Republic, Prague, Czech Republic, 2021.

[29] *Land Survey Office*. Accessed: Apr. 9, 2025. [Online]. Available: https://geoportal.cuzk.cz

[30] *Czech Statistical Office*. Accessed: Apr. 9, 2025. [Online]. Available: https://vdb.czso.cz

[31] Computational and Information Systems Laboratory. (2015). *Degree Global Forecast Grids Historical Archive. Boulder, CO: Research Data Archive At the National Center for Atmospheric Research*. [Online]. Available: https://rda.ucar.edu/datasets/dsd084001/

[32] D. Chaparro, M. Vall-llossera, M. Piles, A. Camps, C. Rudiger, and R. Riera-Tatche, ‘‘Predicting the extent of wildfires using remotely sensed soil moisture and temperature trends,’’ *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2818–2829, Jun. 2016, doi: 10.1109/JSTARS.2016.2571838.

[33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, ‘‘Scikit-learn: Machine learning in Python,’’ *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[34] R Core Team. *A Language and Environment for Statistical Computing, R Foundation for Statistical Computing*. Accessed: Apr. 9, 2025. [Online]. Available: https://www.R-project.org/

[35] *ESRI*. Accessed: Apr. 9, 2025. [Online]. Available: https://www.esri.com/

[36] M. Hubáček, H.-M. Šárka, D. Sládek, K. Dejmal, and K. Šimková, ‘‘Dependence of occurrence of forest fires in Vysocina region on temperature and its spatial distribution,’’ in *Proc. 25th Central Eur. Conf.*, 2018, pp. 281–293.

[37] J. A. Hartigan and M. A. Wong, ‘‘Algorithm AS 136: A K-means clustering algorithm,’’ *Appl. Statist.*, vol. 28, no. 1, p. 100, 1979, doi: 10.2307/2346830.

DAVID SLÁDEK received the degree from the University of Defence, Brno. He has been recognized for his contributions to the field, including a Fulbright Fellowship at the University of Utah and a Helmholtz research fellow position at German Aerospace Center. He is currently a Weather Forecasting Researcher specializing in aeronautical meteorology and data-driven modeling. His research interests include bridging human expertise with machine learning to enhance meteorological interpretation and automation.

FILIP DOHNAL is a Junior Researcher in postdoctoral position at the Department of Societal Security, Faculty of Logistics and Crisis Management, Tomas Bata University, Zlín, Czech Republic. In his research, he focuses on spatial terrain analyses and data gathering for crisis management, public protection, military purpose, and cartography. He mostly focuses on the possibilities of using GIS and its analysis tools for providing significant, relevant and enriching outcomes in geospatial analyses.

FRANTIŠEK PAULUS is currently the Director of the Population Protection Institute (PPI). PPI is a special facility of the Ministry of the Interior—General Directorate of the Fire Rescue Service of the Czech Republic (FRS CR). PPI focuses on the provision of direct and indirect support to the FRS CR. Activities of direct support are emergency activities, specialization activities, education, and information activities. Activities of indirect support are research and development, testing, and standardization. His research and his education work focus on civil protection, civil emergency planning, crisis management, integrated rescue systems, and risk management. In these fields of interest, he has published several books and articles. He also participates in the preparation of strategic and conceptual materials submitted for approval to the Government of Czech Republic, the National Security Council, Committee for Civil Emergency Planning.

JIŘÍ NEUBAUER received the degree from Masaryk University. Czech Republic, in 1998, and the Ph.D. degree in applied mathematics from the University of Ostrava, Czech Republic, in 2006. In 2016, he defended his Associate Professor thesis "Use of quantitative methods in the analysis of defense economy" in the field of "state defense economy" at the University of Defence, Brno, Czech Republic. He is currently the Deputy Head of the Department of Quantitative Methods, University of Defence. His research interests include statistical data analysis, time series analysis, and applications in various fields.

MARTIN MOLEK is currently a Researcher with the Population Protection Institute, a special research and development facility of the Ministry of the Interior–General Directorate of the Fire Rescue Service of the Czech Republic (FRS CR). He focuses his research activities on civil protection, civil emergency planning, crisis management, the integrated rescue system, risk management, and the use of innovative technologies in firefighting. He holds the position of secretary of the certification commission for research and development of FRS CR.

TOMÁŠ ZEMAN received the Ph.D. degree in population protection from the University of Defence, Brno, Czech Republic, in 2015. He is currently an Associate Professor with the Department of Societal Security, Faculty of Logistics and Crisis Management, Tomas Bata University in Zlín, Czech Republic. His current research interests include risk assessment methods and their application in various fields, especially in ensuring the security of soft targets against violent attacks and in human health risk assessment.