



OPEN

Exploration of geo-spatial data and machine learning algorithms for robust wildfire occurrence prediction

Svetlana Illarionova[✉], Dmitrii Shadrin, Fedor Gubanov, Mikhail Shutov, Usman Tasuev, Ksenia Evteeva, Maksim Mironenko & Evgeny Burnaev

Wildfires play a pivotal role in environmental processes and the sustainable development of ecosystems. Timely responses can significantly reduce the damages and consequences caused by their spread. Several critical issues in wildfire behavior analysis include fire occurrence forecasting, early detection, and spread prediction. In this study, we focus on wildfire occurrence forecasting, which is a valuable tool for facilitating earlier intervention. Conventional approaches primarily rely on the computation of fire indices based on weather conditions. However, solutions that utilize more comprehensive environmental data, remote sensing information, and artificial intelligence (AI) algorithms may offer substantial advantages for rapid decision-making and extensive territory monitoring. The wide variety of spatial environmental parameters and the great diversity of geographical regions that influence wildfire occurrence complicate this task. Consequently, there is no unified approach for predicting wildfire occurrences using remote sensing data and AI techniques. The goal of this study is to explore the potential of predicting wildfire occurrences using various available environmental parameters - meteorological, geo-spatial, and anthropogenic - and machine learning (ML) algorithms. We developed a unified pipeline for data acquisition and subsequent ML-based algorithm development. The comprehensive analysis includes the following algorithms: Random Forest, XGBoost, Autoencoder, ConvLSTM, Attention Multilayer Perceptron, and RegNetX. In addition, we explore several metrics to assess the quality of developed models in case of highly imbalanced spatio-temporal data. To conduct the study, we collected a unique dataset covering several large regions in central Russia, incorporating more than 17,000 verified wildfire events over a period of 10 years. The findings underscore the necessity of developing individual ML models tailored to each region, taking into account the specific environmental features correlated with the probability of fire occurrence. The quality of the achieved models, as measured by F1-score, varies from 0.7 to 0.87 depending on the region, demonstrating the potential of integrating such algorithms into emergency response systems.

Keywords Fire spreading, Computer vision, Remote sensing, Deep learning, Classification, Image processing

In recent years, artificial intelligence (AI) algorithms have become a valuable tool in environmental studies, such as ecological condition assessments and forecasting natural hazards or analyzing damage^{1,2}. The task of wildfire prevention is one of the important areas for the application of advanced algorithms, as it aims to minimize risks, save lives, and preserve natural resources. This task can be broadly categorized into two main groups: wildfire detection and prediction³. Detection involves identifying the location and characteristics of fires using data from various sources, such as satellites^{4,5}, drones^{6,7}, and other observational tools^{8,9}. Prediction tasks, on the other hand, focus on estimating the likelihood of fire occurrence based on indirect indicators, or on forecasting the future spread^{10–13} and potential destructiveness of a fire^{14,15} after it has been detected.

In this work, we focus specifically on the challenge of predicting wildfire occurrence over multiple consecutive days. By providing predictions several days in advance, our goal is to enhance early warning systems and support proactive fire management strategies.

Traditionally, fire danger prediction has relied on standardized Fire Danger Rating Systems (FDRS)¹⁶, implemented at state or national levels. Examples include the Canadian (CFFDRS)¹⁷ and Australian (AFDRS)¹⁸

Skolkovo Institute of Science and Technology, Moscow, Russia 121205. email: s.Illarionova@skoltech.ru

Forest Fire Danger Rating Systems, as well as the European Forest Fire Information System (EFFIS)¹⁹. These systems typically employ indices like the Fire Weather Index (FWI)²⁰, Nesterov index²¹ and others for fire risk assessments. Specified indices, calculated from meteorological variables such as temperature, humidity, wind speed, and precipitation, have long served as key tools for evaluating wildfire risk by correlating weather conditions with fire likelihood.

While effective, these traditional methods are limited by their focus on a narrow set of meteorological inputs, often overlooking critical factors like vegetation conditions and human activities. As our understanding of wildfire dynamics has evolved, there is growing recognition of the need for more advanced prediction tools that incorporate a broader range of data.

This is where machine learning (ML) becomes highly relevant. ML is a field of study of artificial intelligence (AI) that focuses on creating systems that learn and evolve based on the data they receive. In the case of the task of daily fire prediction, the data includes both spatial characteristics of the study areas (such as relief and proximity to the sea) and temporal characteristics (such as average daily temperature and wind speed). Many existing works use fire data collected over many years to train an ML model²². Most often, the data is in the form of tabular data containing the date of the burning, the geographic location of the fire, the cause of the burning, and sometimes the resulting damage.

Supervised ML-models based on decision trees, the so-called tree-based models²³ showed high efficiency in the task of fire prediction. Works²⁴⁻²⁶ use Random Forest (RF) algorithm to assess fire hazard in California. The authors use historical data to predict the overall probability of wildfires to produce a map of fire-prone regions. While these studies do not specifically address the problem of daily fire prediction, their fire probability maps can still be instrumental in daily monitoring of target regions. In article²⁷, the authors also solved the fire mapping problem using methods based on decision trees: XGBoost (XG) and RF, respectively, using data from around the world for training. In this work, the XGBoost method showed superiority over RF.

In study²⁸, the authors consider the fire prediction problem as a one-class or two-class classification problem. In a one-class scenario, the model is trained on positive examples (fire events) to define a decision boundary that distinguishes inliers (fires) from outliers (non-fire events) based on a probability threshold. The key challenge in this approach is selecting an accurate threshold. In contrast, two-class classification involves training the model on both fire and non-fire events, allowing it to directly differentiate between the two classes. Their findings suggest that one-class models²⁹, such as One-class SVM (OCSVM), Isolation Forest (IF), and DeepSVDD, outperform two-class models (RF, Logistic Regression (LR), SVM) in the task of predicting fires in California regions. Another approach to supervised learning is demonstrated in works^{30,31}. Both works use genetic algorithms to select the most appropriate function for aggregating input features to obtain an estimate of the probability of fire occurrence.

Recently, deep learning (DL) has emerged as a powerful subset of ML, particularly effective in handling the high dimensionality and complexity of wildfire prediction tasks. Deep learning is a subset of machine learning that uses artificial neural networks (NN) with multiple layers to learn representations of the input data³². A big breakthrough for DL methods was the backpropagation method³³ for training, which, together with the increasing availability of computing power, made DL methods extremely popular.

Multilayer perceptron (MLP) is one of the modern feedforward NN, consisting of fully connected neurons with a nonlinear kind of activation function. MLPs have found wide application in the task of fire forecasting, outperforming traditional machine learning methods in prediction quality^{34,35}.

As a result of the study³⁶, the advantage of convolutional neural network (CNN) models over both the MLP approach and traditional ML methods was shown. The authors developed a convolutional network called AllConvNet to build daily maps of how likely wildfires will occur over the next seven days. Historical data from the Australia region was used to train and test the NN model.

Another alternative to classical ML algorithms and DL algorithms in environmental studies is Convolutional Long Short-Term Memory (ConvLSTM)³⁷, developed in 2015. It is a special CNN for processing time series of images. The architecture of this convolutional neural network is based on the Long Short-Term Memory mechanism and was originally proposed for weather forecasting. The authors of³⁸ used ConvLSTM for daily fire forecasting and showed that this architecture makes better predictions than a baseline CNN model for the task under consideration.

The most existing models are designed to predict fire danger across broad regions, focusing on identifying areas with elevated risk rather than specific fire occurrences. Such predictions, often based on fire danger indices, lack the precision required to pinpoint localized areas at risk. For effective application use, daily fire forecasting systems must predict a more or less accurate location of the fire source in the near future, which is fundamentally different from most fire probability mapping tasks.

Among the works known to us, only a few research teams solve the problem of prediction on a Cartesian grid, convenient for use in automatic fire monitoring systems. Predicting on a fixed grid allows for consistent spatial resolution, better integration with geospatial data, and more efficient resource allocation for fire management. In³⁹, authors handle the problem of next day fire prediction using dataset comprises a geographic grid of high granularity (each cell being 500m wide) covering the whole Greek territory. The study demonstrates the superiority of the MLP approach over some classical approaches (RF, XG, LR), but also highlights that for the grid prediction task, CNN⁴⁰ approaches are likely to be effective.

Besides the various AI algorithms that can be implemented for wildfire forecasting, optimal data source selection is a key component for advanced GIS development capable of real-time performance. Often, the data for different environmental parameters vary in their spatio-temporal properties. Therefore, the general pipeline for effective solution creation should involve data selection, processing, and fusion. Relevant environmental features that can affect wildfire occurrence should be discussed in detail.

In this study, we explore the task of wildfire prediction over a 5-day period using openly available environmental data. Although there are existing works on wildfire prediction, they typically focus on a single type of AI algorithm, such as classical ML algorithms, deep learning (DL) algorithms, or anomaly detection models. Moreover, these studies usually involve a small territory with relatively uniform environmental conditions, which does not allow for a comprehensive assessment of the capabilities of the chosen algorithms. To address this challenge, we evaluate the most relevant AI algorithms that can be adapted to the problem at hand and propose a pipeline for data collection and fusion for AI algorithm training. Our study includes experiments with Random Forest, XGBoost, Autoencoder, ConvLSTM, Attention Multilayer Perceptron, and RegNetX models. These algorithms represent our problem statement by leveraging either tabular or image data. In addition to data fusion and model development, we investigate the topic of fair and representative model quality assessment in the context of highly imbalanced spatio-temporal data on wildfire occurrences under natural conditions. For a comprehensive study involving varying environmental conditions, we collected a dataset comprising more than 17,000 wildfires across four large regions of Russia over a span of 10 years. Overall, the goal of this study is to shed more light on the challenging task of natural hazard forecasting in the absence of a currently unified pipeline and to emphasize the wide range of geo-spatial parameters that should be considered and analyzed in depth. The main contributions are the following:

- We proposed a methodology of environmental data collection and processing for wildfire occurrence prediction based on openly-available and regularly updated data sources;
- We developed and adapted several scenarios of various ML algorithms application for wildfire prediction;
- We proposed and compared various metrics and sampling techniques for comprehensive analysis of the developed algorithms;
- We conducted detailed feature analysis for deeper understanding of model behavior in various environmental conditions.

The general workflow of this paper is structured to systematically address the challenge of wildfire occurrence prediction through a series of well-defined steps, each corresponding to a specific chapter. In the **Methodology and data** chapter, we describe the collection and preprocessing of a comprehensive dataset, including meteorological, geo-spatial, and anthropogenic features, as well as the clustering of fire events and balanced sampling to handle class imbalance. The **Algorithms** chapter focuses on the development and adaptation of various machine learning models, including classical methods such as Random Forest and XGBoost, deep learning approaches such as ConvLSTM and RegNetX, and anomaly detection techniques like Autoencoder, tailored for region-specific fire prediction. In the **Results** chapter, we evaluate the performance of these models using a range of metrics, including F1-score and custom balanced metrics, and provide visualizations of predicted fire probabilities to assess spatial accuracy. Finally, the **Discussion** chapter analyzes the importance of different features across regions, explores the distribution of meteorological data relative to fire events, and discusses the implications of our findings for future wildfire prediction systems.

Methodology and data

Problem statement

The goal of daily fire forecasting is to use previously observed data sequences to predict a fire occurrence in a local region. Suppose we are observing a dynamic system in a spatial domain represented by a set of points $\{x\}$ where the observation is carried out. Each point contains measurements in its surroundings that change or do not change over time (temporal or non-temporal parameters), these could be some weather measurements, landscape shape parameters, human population density statistics, vegetation indexes, and many other factors. An observation in region x on day t can be represented by a feature tensor $\phi_t(x)$. If we periodically record observations, we obtain a sequence of tensors $\{\phi_{t-J+1}, \phi_{t-J+2}, \dots, \phi_t\}$. Thus, the spatio-temporal fires forecasting problem is to predict the most likely k -length sequence in the future given the previous J observations:

$$\tilde{f}_{t+1}(x), \dots, \tilde{f}_{t+k}(x) = \arg \max_{f_{t+1}(x), \dots, f_{t+K}(x)} p(f_{t+1}(x), \dots, f_{t+K}(x) | \phi_{t-J+1}(x), \phi_{t-J+2}(x), \dots, \phi_t(x)) \quad (1)$$

where k is the number of future days for which you need to make a prediction about fire, J is the number of previous days used to generate the prediction, and

$$f_t(x) = \begin{cases} 0, & \text{if there is no fire in } (x) \text{ at day } t \\ 1, & \text{otherwise} \end{cases}$$

is a function of prediction for future days whether a fire will or will not occur;

Often the function $f_t(x)$ is replaced by the function of predicting the probability of a fire occurring in a particular area. The daily fire forecasting problem naturally becomes a spatiotemporal sequence forecasting problem. One of the boundary cases of fire prediction is a prediction only for the next day, then $k = 1$. Another boundary case is a prediction based not on a sequence of days, but only on one day ($J = 1$) or data aggregated over several previous days.

In our study, instead of focusing on predicting daily fire occurrences for each of the following k days, we aim to forecast the probability of a fire occurring within any of the next K days. We consider a length of 5 days ($K = 5$) because it is a suitable forecasting horizon from a practical standpoint and allows the model to handle the uncertainty of fire occurrence events. By predicting the probability over this 5-day period, it is possible to localize areas prone to wildfire occurrence due to environmental conditions. Firefighters can then take measures to prevent or minimize such risks. This approach simplifies the temporal forecasting problem into a single

probabilistic prediction, capturing the likelihood of fire occurrence over a given future time window, rather than specific daily outcomes.

Formally, our goal is to find the binary function that most likely describes the occurrence of a fire at location x within the time interval $[t+1, t+K]$, given the sequence of past observations:

$$\tilde{f}_{[t+1:t+K]}(x) = \arg \max_{\bigcup_{\tau=t+1}^{t+K} \{f_{\tau}(x)\}} p \left(\bigcup_{\tau=t+1}^{t+K} \{f_{\tau}(x)\} \mid \phi_{t-J+1}(x), \phi_{t-J+2}(x), \dots, \phi_t(x) \right) \quad (2)$$

Here, the function $\tilde{f}_{[t+1:t+K]}(x)$ represents the most probable binary outcome (whether a fire will or will not occur) in the region x within the next K days, based on the observed data sequence up to time t . In the case of predicting probabilities, (2) turns into (3):

$$\tilde{p}_{[t+1:t+N]}(x) = \max_{F(x, K)} p(F(x, K) \mid \phi_{t-J+1}(x), \phi_{t-J+2}(x), \dots, \phi_t(x)) \quad (3)$$

where $F(x, K)$ replaces the union of functions in the previous expression and describes the probability of at least one fire occurring in the next N days.

Obviously, (2) is completely the same as statement (1) in the case where $K = 1$, but when $K >= 2$, statement (2) has a different meaning. It is worth noting that from the set of values of $f_t(x)$ we can obtain $F(x, K)$, but the opposite is impossible in general. However, this probabilistic approach still allows for a flexible and aggregated assessment of fire risk, which is particularly useful for operational decision-making and resource allocation in fire management.

Study area

The research area covers 4 administrative regions (oblast) of the Russian Federation: Amur, Irkutsk, Rostov, and Sverdlovsk, as shown in Fig. 1.

Amur Oblast is located on the banks of the Amur and Zeya rivers in the Russian Far East. It has two different climates and is dominated by monsoon-influenced subarctic climate⁴¹. The region has a population of approximately 750 000⁴² and covers a total area of 361 900 km²⁴³. Average temperatures in January range from -23.5 to -21.8°C , while July temperatures are from $+21.2^{\circ}\text{C}$ to $+18^{\circ}\text{C}$. The average annual precipitation is around 674 millimeters⁴¹.

Irkutsk Oblast, located in southeastern Siberia in the basins of the Angara, Lena, and Nizhnyaya Tunguska Rivers, characterized by subarctic climate. The population is about 2 330 000⁴², and its total area is 774 800 km²⁴³. The average temperatures in January vary from -20.6 to -19.6°C , and in July from $+18.1^{\circ}\text{C}$ to $+20^{\circ}\text{C}$. The Average annual precipitation is approximately 452 millimeters⁴⁴.

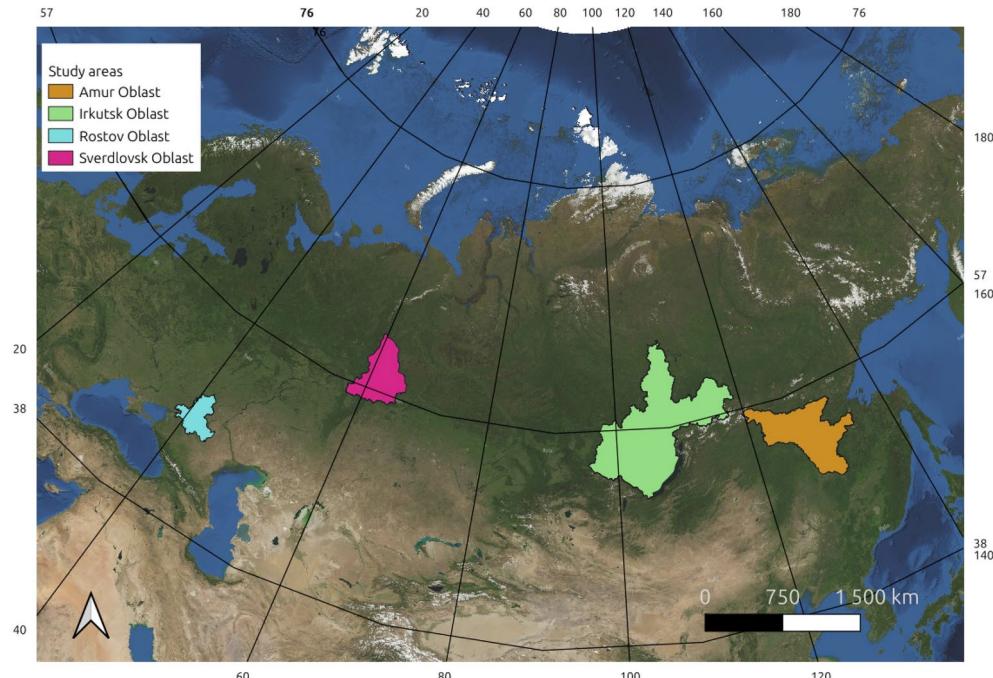


Fig. 1. Study area comprises 4 regions from left to right: Rostov Oblast, Sverdlovsk Oblast, Irkutsk Oblast, Amur Oblast. The figure is created by the authors using QGIS v.3.22 software (<https://qgis.org/en/site/>), Yandex Satellite composite derived from QGIS plugin QuickMapServices (<http://qms.nextgis.com>) is chosen for visualization.

Rostov oblast is situated in the Pontic-Caspian steppe, directly north over the North Caucasus and west of the Yergeni hills. It has a population of 4,150,000⁴² and an area of 101 000 km²⁴³. Region has a hot humid continental climate. Average January temperatures range from -3.5°C to 1.9°C , and in July temperatures vary from $+24.2^{\circ}\text{C}$ to $+24.9^{\circ}\text{C}$. The average annual precipitation is about 460 millimeters⁴⁵.

Sverdlovsk Oblast, located in the eastern slopes of the Middle and North Urals and the Western Siberian Plain, has a population of 4 230 000⁴² and total area of about 194 300 km²⁴³. This region is dominated by a moderately continental climate. The average January temperatures vary from -14.7°C to 14.3°C , and July temperatures range from $+17.9^{\circ}\text{C}$ to $+19.2^{\circ}\text{C}$. The average annual precipitation is about 601 millimeters⁴⁶.

The Fig. 2 displays the distribution of ignition points across described regions. The number of ignition points varies significantly. Irkutsk has the highest number of ignition points, with 55,189, followed by Amur with 40,770. Rostov and Sverdlovsk have considerably fewer ignition points, with 4991 and 3010, respectively. The data confirm that large regions of Eastern Siberia are highly susceptible to forest fires.

Reference data

Fire points are thermal anomalies identified by the results of satellite imagery after thematic processing. The thermal cause may be the burning of garbage, a man-made process, or forest wildfire. In total, 105 thousand fire points manually verified as a category of forest wildfires were considered in 4 selected regions for 2012–2022. They are uniquely described by coordinates and date. In order to work with fire events, it is necessary to define for each fire point which fire it belongs to. For this purpose, clustering on spatial and temporal axes is used. The clustering process is described in detail in 'Data preprocessing' section of this article.

Remote sensing and geospatial data

In this study, we collected dataset involving 14 environmental variables such as topography, population, remotely sensed, and climatic data from March 1st to October 31st for 2012–2022 years.

A topography-related variables, such as elevation, aspect and slope, have effect on local climate and vegetation types, amount and intensity of solar radiation received by a given location, human accessibility^{47,48}. Elevation data was retrieved from the Copernicus GLO-90 Digital Elevation Model⁴⁹ with 90 m spatial resolution. Aspect and slope variables were derived through processing techniques applied to the elevation data, with a detailed explanation available in 'Data preprocessing' section of this article.

Higher population densities significantly impact fire occurrence^{48,50,51}, as they are often associated with increased human activity in residential, infrastructural, and recreational zones. Common activities such as campfires, outdoor burning, discarded cigarettes, and equipment use, can act as ignition sources for wildfires. We obtained population density data from the WorldPop dataset⁵² with a spatial resolution of 30 arc-seconds (approximately 1 km at the equator).

The remote sensing variables selected for fire occurrence prediction were collected using the Application for Extracting and Exploring Analysis Ready Samples⁵³ and included the following products in 500 m spatial resolution:

1. Land cover from MCD12Q1 v061⁵⁴ dataset. Different land cover types provide varying amounts and types of fuel for wildfires. For example, dense forests typically have abundant vegetation that can serve as fuel for fires, while grasslands may have shorter, more easily ignitable vegetation. Land cover classifications also provide insights into human land use and development patterns, which can influence fire occurrence. Urban and developed areas may have reduced vegetation cover and fuel availability compared to natural or rural areas. However, human activities in urban and peri-urban areas, such as construction, vehicle use, and outdoor recreation, can still pose fire risks.

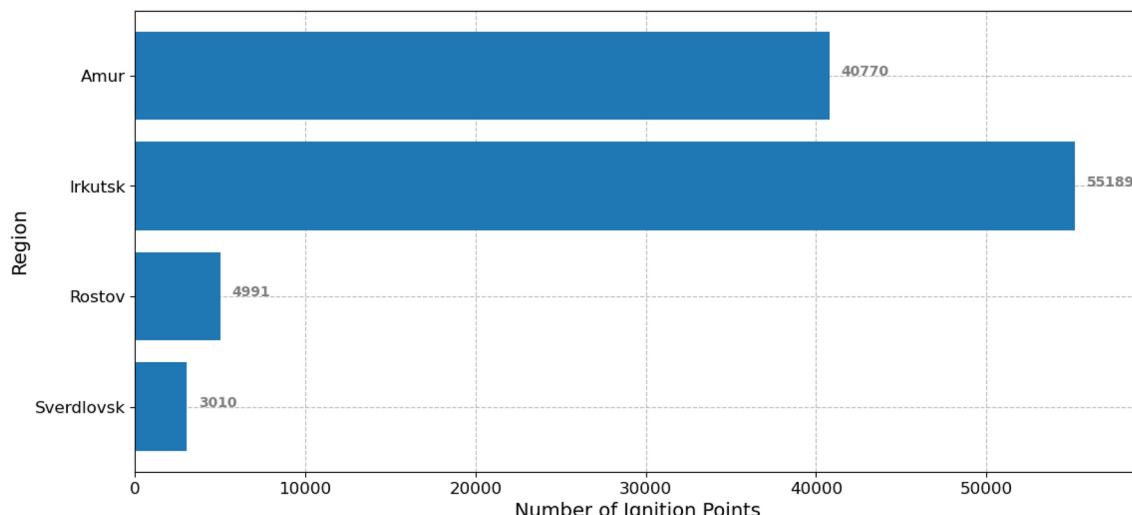


Fig. 2. Distribution of ignition points by regions of study area.

- Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) from MOD13A1 v061⁵⁵ dataset. NDVI and EVI are both widely used remote sensing indices that provide valuable information about vegetation health and density. Higher NDVI and EVI values typically indicate denser vegetation, which can serve as fuel for wildfires⁵⁶.
- Evapotranspiration (ET) and Potential ET (PET) from MOD16A2GF v061⁵⁷ dataset. ET measures the amount of water transpired by plants and evaporated from the soil surface. Low ET values indicate drier conditions, potentially leading to increased vegetation stress and higher fire risk. PET represents the maximum amount of water that could be evaporated from the soil and transpired by vegetation under prevailing environmental conditions. PET influences vegetation moisture stress, with higher PET values indicating greater water demand and potential vegetation desiccation.
- Fraction of Photosynthetically Active Radiation (FPAR) and Leaf Area Index (LAI) from MOD15A2H v061⁵⁸ dataset. FPAR quantifies the fraction of incoming solar radiation absorbed by vegetation canopy⁵⁹. High FPAR values indicate active vegetation growth and biomass accumulation, which can contribute to increased fuel loads and fire risk⁵⁶. LAI measures the total area of leaves per unit ground surface area⁶⁰. High LAI values suggest dense vegetation cover and greater fuel continuity.

Meteorological conditions, such as temperature, wind speed, and precipitation, are widely recognized as key determinants of forest fire occurrence⁴⁷. 6 climate-related variables – air and dewpoint temperature at 2 m above the surface, 10m u-component and v-component of wind and total precipitation – were retrieved from ERA5-Land dataset at The Climate Data Store⁶¹. The spatial resolution of the obtained variables was 0.1 degree with a 3 hour temporal resolution starting from 00:00 UTC.

Table 1 provides a summary of the initial data characteristics, prior to undergoing any preprocessing procedures.

In addition to the aforementioned attributes, we incorporated two additional variables: the day number of the year and the geographical coordinates (latitude and longitude) of the forecast area. Including the day number helps capture seasonal variations, as the likelihood of fire occurrence is often closely tied to specific times of the year due to factors like vegetation cycles, temperature changes, and periods of drought. The geographical coordinates allow the model to account for the spatial heterogeneity of fire risks. By explicitly introducing these features, we aim to better capture the dependencies between fire probability, seasonal timing, and location, thus enhancing the model's ability to generalize across different regions and time periods.

The selected set of features represents a widely accepted combination for wildfire prediction^{22,27,36,38}. Topographic variables are inherently independent of other features, as they describe static terrain characteristics. Population density is a key indicator of human activity, separate from other physical environmental variables. Meteorological variables are the main source of information for capturing dynamic environmental conditions that are critical for predicting wildfire occurrence. While land cover may correlate with vegetation indices, the latter provide essential information about current vegetation state, which directly influences fuel availability. FPAR and LAI have a known relationship with NDVI and EVI, but they offer complementary insights into photosynthetic activity and vegetation structure. Potential correlations between ET, PET, and total precipitation reflect their shared dependence on moisture availability. Notably, ET and PET, derived from satellite data, complement meteorological variables by providing insights into soil moisture and drought stress, key indicators of fire-prone conditions. Selected set of features provide sufficient coverage of the essential static and dynamic drivers of wildfire occurrence. Although further optimization may be explored, the current feature set avoids significant redundancy that could impact model performance while maintaining interpretability and predictive reliability.

Variable	Spatial resolution	Temporal resolution	Scale factor	Valid range	Fill value	Units
Elevation	90 m	Single composite representing data collected between 1 January 2011 and 1 July 2015	1	NA	NA	m
Population density	30 arc-seconds	Yearly from 2000 to 2020	1	NA	-99999	people/km2
Land cover	500 m	Yearly from 1 January 2001 to 31 December 2022	1	[1, 17]	255	Classes
EVI, NDVI	500 m	Composite from the 16 day period (best pixel value) from 18 February 2000 to present	0.0001	[- 2000, 10000]	- 3000	Dimensionless
FPAR, LAI	500 m	Composite from the 8 day period (best pixel value) from 18 February 2000 to present	0.01 for FPAR, 0.1 for LAI	[0, 100]	[249, 255]	Dimensionless
ET, PET	500 m	Composite from the 8 day period (sum) from 1 January 2000 to present	0.1	[- 32767, 32700]	[32761, 32767]	kg/m2/8d
2m dewpoint and air temperature	0.1 degree	Hourly from 1950 to present	1	NA	NA	K
10m u- and v-component of wind	0.1 degree	Hourly from 1950 to present	1	NA	NA	m/s
Total precipitation	0.1 degree	Hourly from 1950 to present	1	NA	NA	m

Table 1. Characteristics of openly available data.

Data preprocessing

For all collected data, series of preprocessing steps were implemented to ensure consistency in spatial extent and resolution. Initially, the data were cropped to the region boundaries delineated in the section '[Study area](#)', ensuring that only relevant geographic areas were retained for analysis. Subsequently, we standardized the spatial resolution of the datasets to 0.0059435 degree per pixel (approximately 650 meters). To achieve both of these preprocessing steps, we utilized the gdal.Warp function of Python osgeo library⁶² with the bilinear resampling algorithm.

Furthermore, for remote sensing data, an additional preprocessing step involved scaling the data by a specified factor and assigning a fill value to all values falling outside a valid range. Scale factors, valid ranges, and fill values are specified in the Table 1.

In the case of elevation data, specific processing methods were applied to derive additional topographic variables. We utilized the osgeo.gdal.DEMProcessing function to compute both aspect and slope from the elevation. The 'aspect' mode with the 'zeroForFlat' attribute set to True was utilized to calculate aspect. Similarly, the 'slope' mode with 'slopeFormat' parameter set to 'degree', 'scale' parameter set to 111120 (used for calculations in meters), and 'computeEdges' set to True, was employed to compute slope values.

Weather data preprocessing included, in addition to the steps described above (cropping, converting to a single resolution), also aggregation of meteorological attributes values for the day and calculation of the Nesterov index. The aggregation methods differed depending on the meteorological attribute: precipitation - summation, temperature - mean and maximum, for the other attributes the mean value was calculated. The following formula was used to calculate the Nesterov index:

$$NI_i = NI_{i-1} * K_i^P + t_i * (t_i - t_i^D), \quad K_i^P = \begin{cases} 1, & \text{if precipitation was less than 3mm} \\ 0, & \text{otherwise} \end{cases}$$

where NI_i is Nesterov index value for i -th day, K_i^P is coefficient of precipitation corrections on i -th day, t_i is max temperature on i -th day and t_i^D is mean dew point temperature on i -th day.

Figure 3 presents preprocessed population density, elevation, slope, aspect, and land cover data for the year 2020 within a sample area of the Irkutsk Region. Figure 4 represents preprocessed remote sensing data, and Fig. 5 depicts weather data for the same sample on 11 July 2020.

We chose minimax normalization as one of the stages of data preprocessing. For the entire available data volume, 0.001 and 0.999 quantiles were calculated for each feature for subsequent minimax normalization.

The total number of collected features for each fire is 56. The list of all features are the following:

1. Topography features: elevation, aspect and slope;
2. Weather features: 6 daily measurements (e.g., temperature, total precipitation) collected for each of the 7 days leading up to the fire;
3. MODIS features: Land cover, EVI/NDVI, FPAR/LAI, ET/PET;
4. Population density feature;
5. Additional features: day number of the year and coordinates (latitude, longitude).

Fire points preprocessing

Initially, we only had data on verified fire points, where each fire point is characterised by coordinates and date. However, several fire points can belong to the same fire event. Clustering along the spatial and temporal axes is required to partition fire points into independent fire events. A static 0.2-degree grid is used to cluster the fire points in space, which is defined for each study region. In the next step, we assign fire points to the same fire in case they lie in the same cell and the time difference does not exceed 6 days. In this way, we break down all fire points into independent fire events. The clustered ignition points for the sequence of days are shown in Fig. 6.

After clustering, the number of fires and ignition points for each region is as follows: Amur - 7042 fires from 40,770 points, Irkutsk - 7733 fires from 55,189 points, Rostov - 2664 fires from 4,991 points, and Sverdlovsk - 1139 fires from 3010 points. In some regions, such as Amur, a single fire corresponds to an average of 5.8 ignition points, while in others, like Rostov, the ratio is much lower, with approximately 1.9 ignition points per fire. This discrepancy may be related to the duration of fires, as longer-lasting fires tend to generate more ignition points. The total number of fires is approximately 17 thousands.

Sampling

We call one sample a data set (weather data, population data, vegetation indices, etc.) for one 0.2×0.2 degree cell (similar to the grid used in [Fire points preprocessing](#)) for a given day. Fire samples are examples in which there was a fire in the selected cell on the corresponding day. Accordingly, non-fire samples are examples where there was no fire in the selected cell on the corresponding day. To form a set of fire samples, we used clustered fire points data (see the [Fire points preprocessing](#) section). To form a set of non-fire samples, we used the following algorithm for each cell of the selected area:

1. Formation of all potential samples. For each cell, we form samples for each day of the given time interval. At this stage, the number of potential samples is equal to $N_{cells} * N_{days}$, where N_{cells} is the number of grid cells in the selected area, N_{days} is the number of days in the specified time interval.
2. Removing samples with possible fires. We remove samples with potential fires from the generated set. To do this, we use data on fire points: for each fire point, we find a cell, and for this cell, we remove the sample for the date the fire was observed, as well as samples for the week ahead and the week before (see Fig. 7).

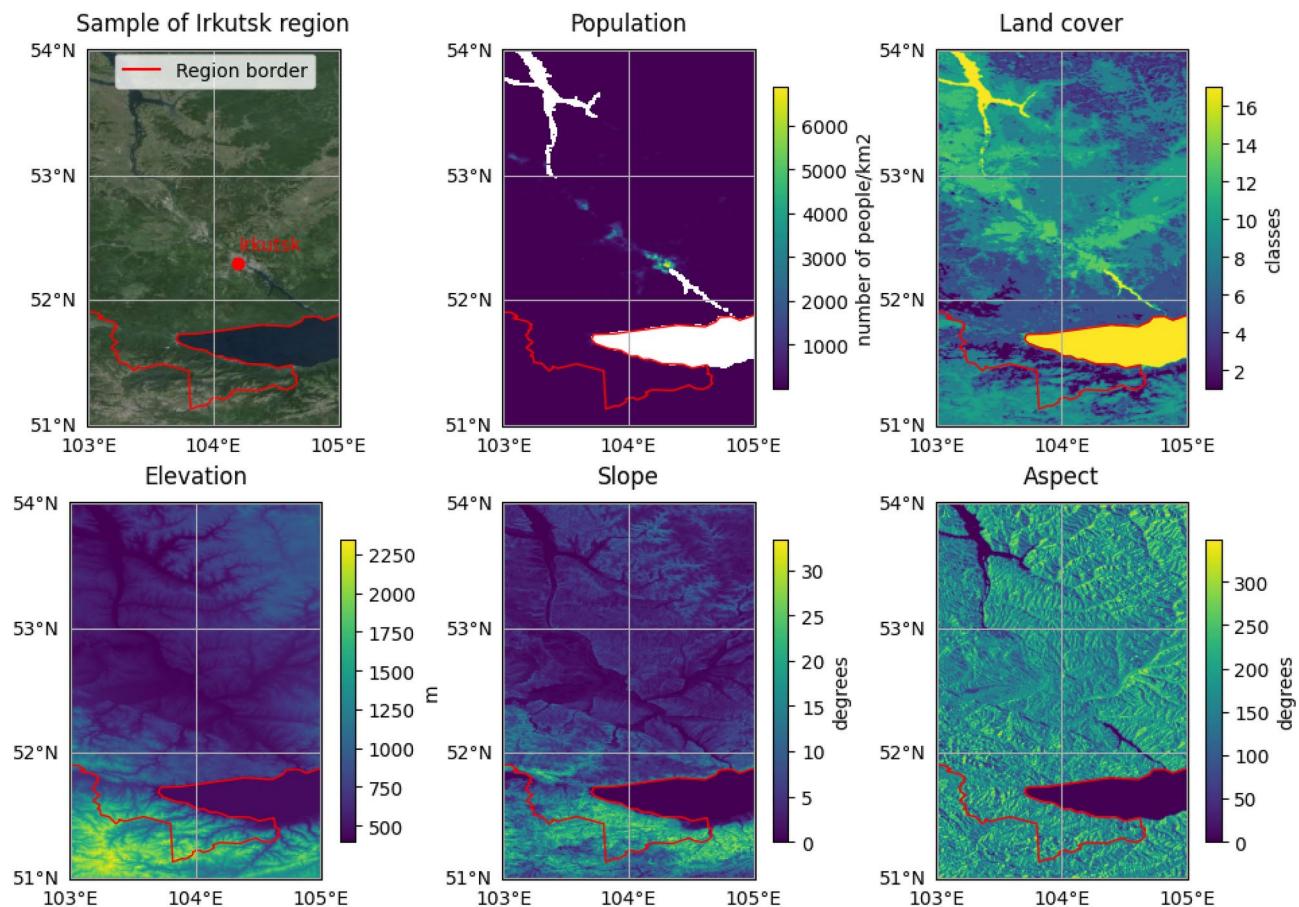


Fig. 3. Preprocessed population density, elevation, slope, aspect, and land cover data for the year 2020 within a sample area of the Irkutsk Region. The map in the upper left corner was generated with the QGIS v.3.22 software (<https://qgis.org/en/site/>) and RGB satellite composite from Google Maps layers available in QGIS.

3. Random sampling. From the obtained set of samples, we randomly select the required number of samples. To address potential seasonal overfitting and maintain a balanced representation of samples with and without fire, we increased the likelihood of selecting samples from months with a large number of fires. Additionally, to mitigate spatial overfitting, we imposed a limit on the number of samples taken from the same location. We also used data augmentation techniques, such as rotation and elastic transformations⁶³, to further enhance the diversity of our sample set.

Data mining

The largest of the regions considered is Irkutsk Region. For this region, unloading and preprocessing all the features in 10 years is an extremely computationally expensive task, and also requires huge storage volumes. It was decided to unload the subdomain of this region, which contains at least 70% of fires in 2012-2021. Thus, we can speed up preprocessing several times and reduce the memory required for storage, sacrificing a small number of fires. Figure 8 demonstrates the subdomain found that contains at least 70% of the fires for Irkutsk Region.

Datasets

We collected datasets on wildfires from 2012 to 2022 for each studied region. Alongside wildfire samples, the datasets include numerous cases of non-fire samples (see the [Sampling](#) section). We aimed to preserve the natural class imbalance inherent to the task, while avoiding excessive imbalance. Wildfire data from the same year was restricted to a single subset (training, validation, or test) to prevent target leakage. Table 2 provides a detailed breakdown of dataset distribution across regions and subsets. To enhance model training, we employed augmentation techniques, allowing us to expand the number of examples in the training set. As a result, the final distribution of samples in the training set may differ from the original dataset.

Algorithms

For the prediction problem, we applied and compared three types of ML approaches: classical ML algorithms for tabular data processing, DL algorithms for image data processing, and anomaly detection methods. The selection of models was guided by the specificity of the data and the requirements of the task. Classical ML algorithms

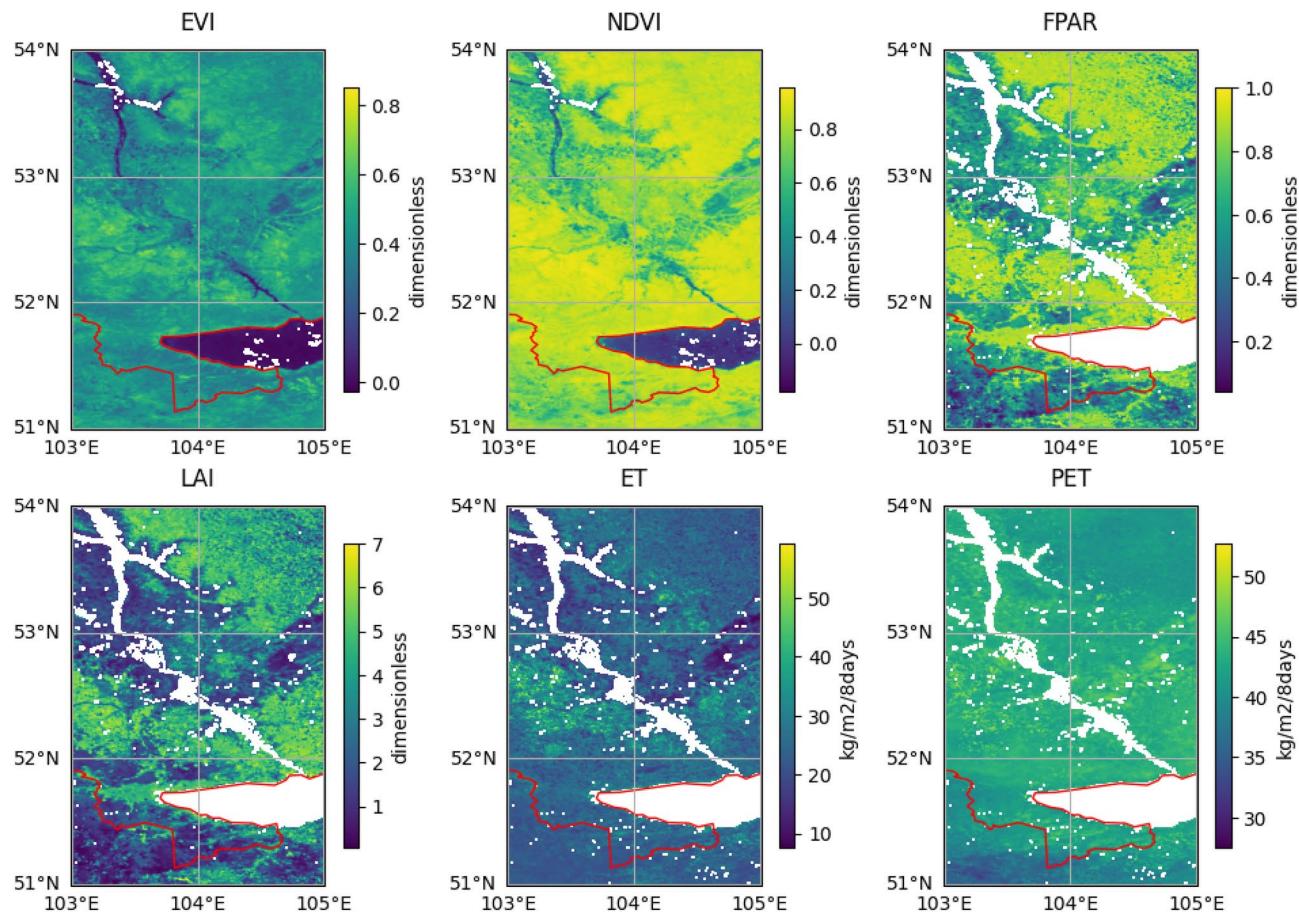


Fig. 4. Preprocessed remote sensing data within the sample area of the Irkutsk Region for 11 July 2020.

were included for their proven effectiveness in handling heterogeneous features^{64,65}. DL models were selected to evaluate their ability to capture both spatial and temporal dependencies. As a robust baseline, we chose RegNetX⁶⁶, an efficient convolutional neural network known for its strong performance in tasks that involve image recognition. Furthermore, we tested ConvLSTM, a state-of-the-art model for predicting spatio-temporal features, especially in the context of natural phenomena forecasting^{37,67–69}. Alongside these established models, we introduce and validate AMLP (Attention MLP), which enhances the ability of standard MLPs to process complex data structures, drawing inspiration from transformer-based architectures⁷⁰, which has demonstrated exceptional performance in various time-series forecasting tasks. Anomaly detection algorithms, including MLP Autoencoder, were also explored to identify fire occurrences by modeling deviations from normal patterns. These diverse approaches allowed us to assess the strengths and limitations of various techniques for wildfire prediction.

1. Classical ML algorithms

Random forest (RF)⁶⁴ is one of the most popular algorithms used to predict fires. RF is an ensemble ML technique used for both classification and regression analysis. It applies both the technique of bagging which is a method of generating a new dataset from an existing dataset and a decision tree concept. We used an RF implementation from an open source python library scikit-learn⁷¹. To find the best estimator, we also varied the IF *n_estimators* parameter from 50 to 500.

XGBoost (XG), Gradient Boosting Decision Trees (GBDT),⁶⁵ is a decision tree ensemble learning algorithm similar to RF, for classification and regression. The process of additively generating weak models is formalized as a gradient descent algorithm over an objective function. We trained XG similarly to RF. We used an open source XG implementation⁷²;

2. DL algorithms

RegNetX. The RegNetX design is straightforward, it consists of a simple stem - initial convolutional layers (stride-2, 3×3 conv. with $w_0 = 32$ output channels), followed by the network body that performs the bulk

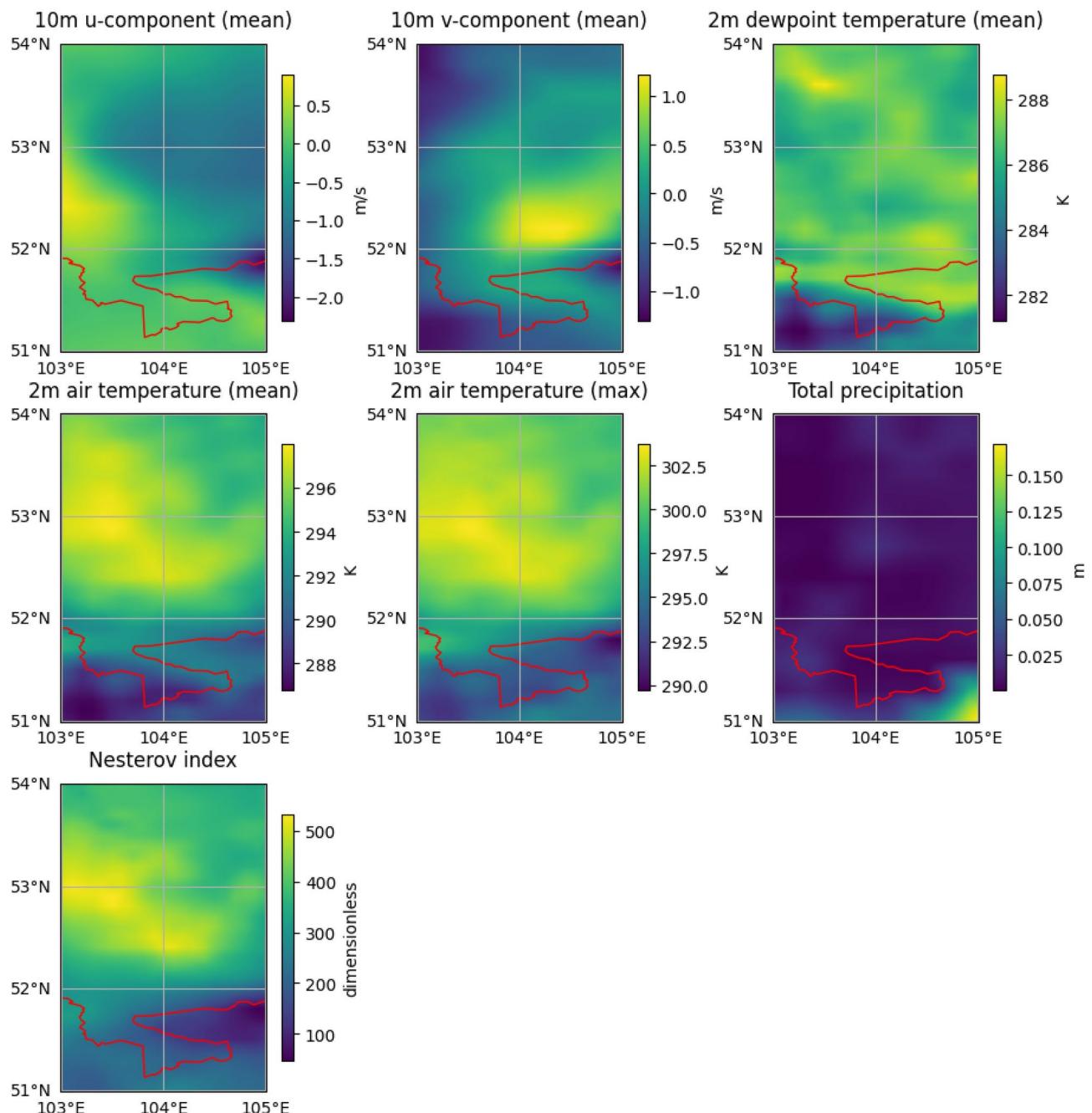


Fig. 5. Preprocessed weather data within the sample area of the Irkutsk Region for 11 July 2020.

of the computation, and a final network head. The network's body is organized into stages, where each stage might operate at a different resolution and feature depth. Every stage is composed of building blocks that follow a standardized design, the blocks contain common layers: (grouped) convolution, batch normalization, ReLU activation and skip connections (see Fig. 9). We used original implementation of *RegNetX_002* backbone without pretraining⁶⁶. The last stage of the network has been trimmed to make the architecture lighter, thus it was possible to reduce the number of trainable parameters by approximately 80%. Additionally we adopted head layers as shown in Table 3;

Attention MLP (AMLP). The main building block of the model – 'attention_mlp' block, is shown in Fig. 10. The design of this block is close to the 'Gated MLP' block described in the article⁷⁰ was adapted for current task. The architecture of this model consists of two parts - encoder and predictor (Fig. 10). Encoder consists of 'attention_mlp' blocks sequence, and decoder part has sequence of blocks containing a Linear, Activation, and a Dropout layers.

ConvLSTM. We conducted additional experiments with recurrent neural network due to the fact that the meteorological attributes used are naturally represented as a spatial-temporal series. And the dynamics of changes

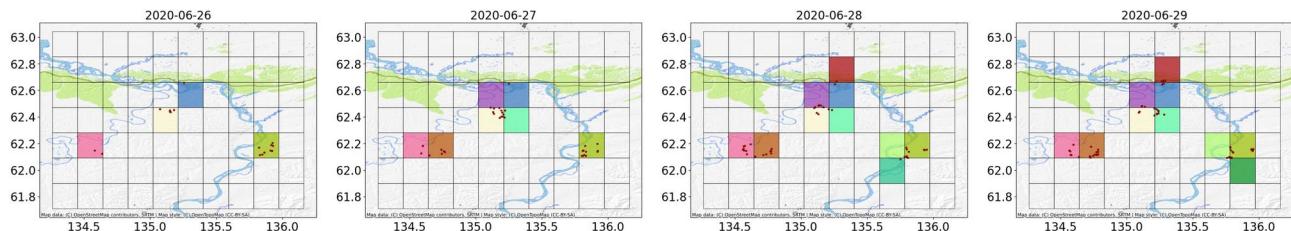


Fig. 6. Example of fire point clustering in the central part of the Republic of Sakha. The sequence of four days with the corresponding fire points (red dots) is shown. The color of the cell corresponds to a unique fire identified through clustering. The figure is created by the authors using Matplotlib library version 3.5.3 (<https://matplotlib.org/>) on Python 3.10 version, OpenTopoMap composite derived from OpenStreetMap (<https://opentopomap.org/about>) is chosen for visualization.

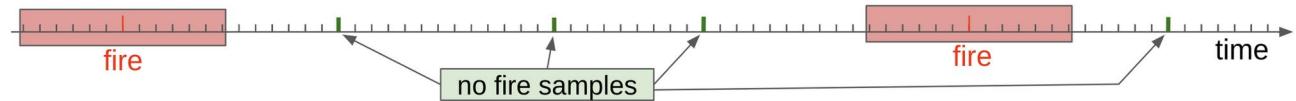


Fig. 7. Method for sampling non-fire samples. Red areas indicate time intervals that are not used to generate examples without fires.

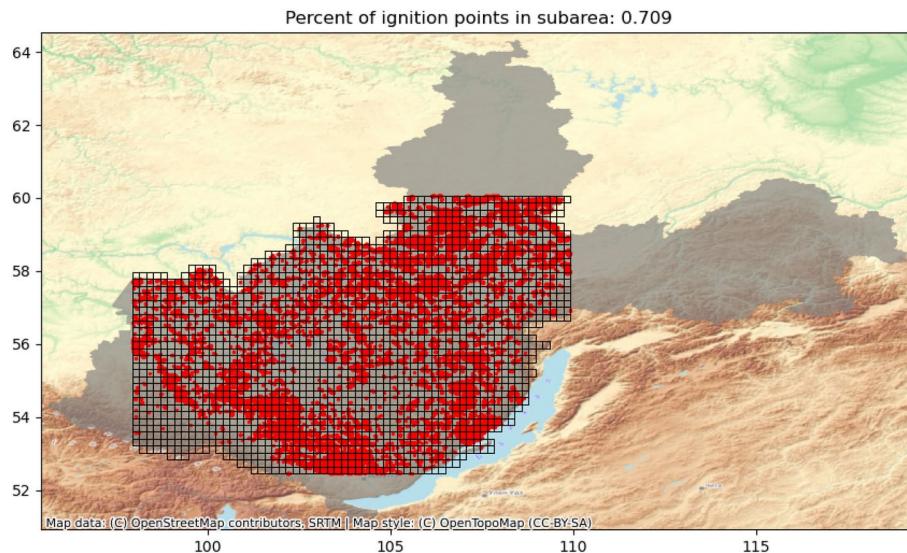


Fig. 8. Subdomain containing at least 70% of the fires for Irkutsk Region. The grey colour represents the full area of the region, the red colour shows the fires that fell into the subdomain. The figure is created by the authors using Matplotlib library version 3.5.3 (<https://matplotlib.org/>) on Python 3.10 version, OpenTopoMap composite derived from OpenStreetMap (<https://opentopomap.org/about>) is chosen for visualization.

Region	Train		Validation		Test	
	Fire	No fire	Fire	No fire	Fire	No fire
Amur	5500	35,445	961	3710	581	3993
Irkutsk	6776	44,830	254	4996	703	5054
Rostov	2325	12,346	215	1462	124	1436
Sverdlovsk	1036	9787	71	1097	32	1141

Table 2. Datasets composition details across study regions.

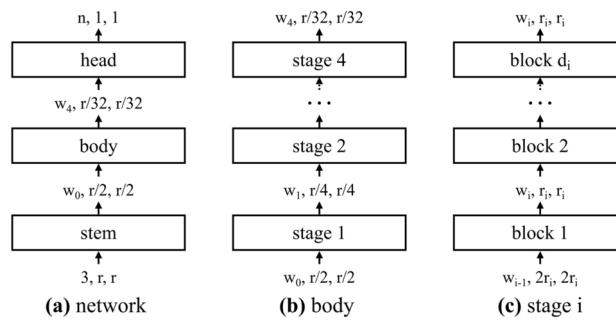


Fig. 9. General network structure for RegNetX models.

Layer	Output shape	Parameters
RegNetX_002 (trimmed)	(BS, 152, 2, 2)	260,752
Conv2D(1, 1)	(BS, 76, 2, 2)	11,628
Conv2D(2, 2)	(BS, 76, 1, 1)	92,492
Conv2D(1, 1)	(BS, 152, 1, 1)	11,704
ReLU + BatchNorm	(BS, 152, 1, 1)	304
Linear	(BS, 1)	153

Table 3. Details of RegNetX architecture. RegNetX Encoder + Head Layers, BS - minibatch size.

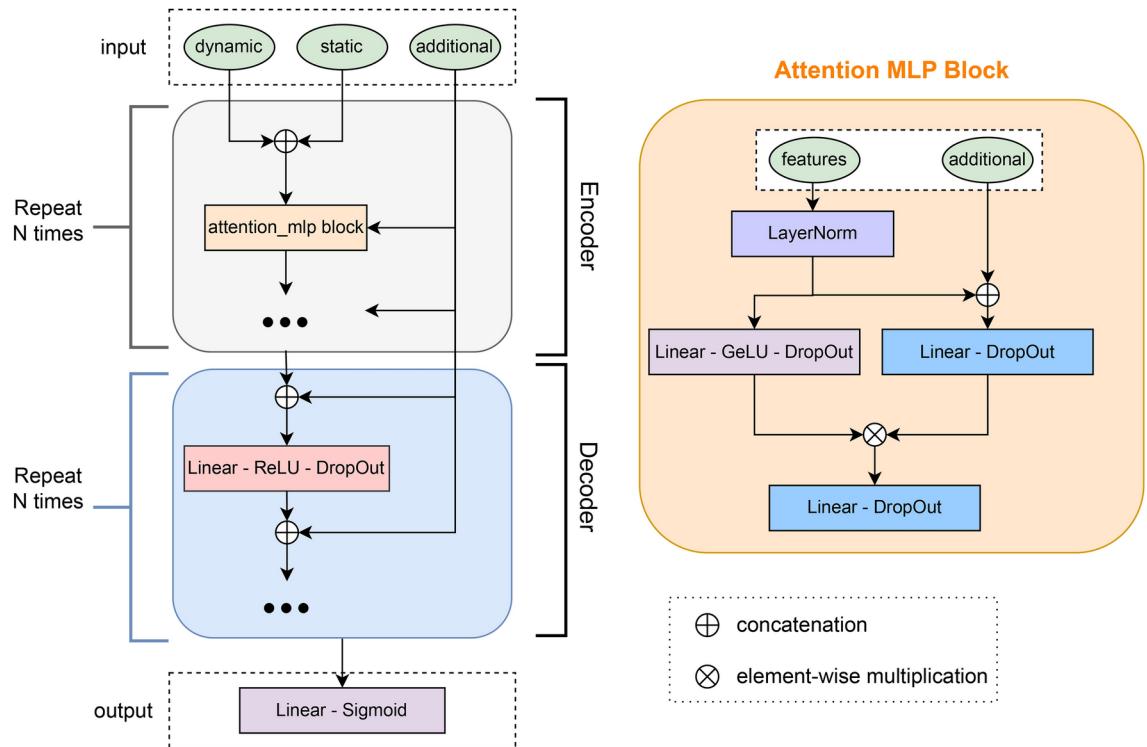


Fig. 10. Attention MLP network. Left side is general network structure, right side is attention MLP block structure.

in meteorological attributes provides important information for fire forecasting. The basic block of this model is ConvLSTM cell, its detailed description is given in the article³⁷. We used original implementation of ConvLSTM backbone and adapted it to our task (Fig. 11). Adaptation of the architecture is necessary primarily to fit the method to the task at hand and consists of:

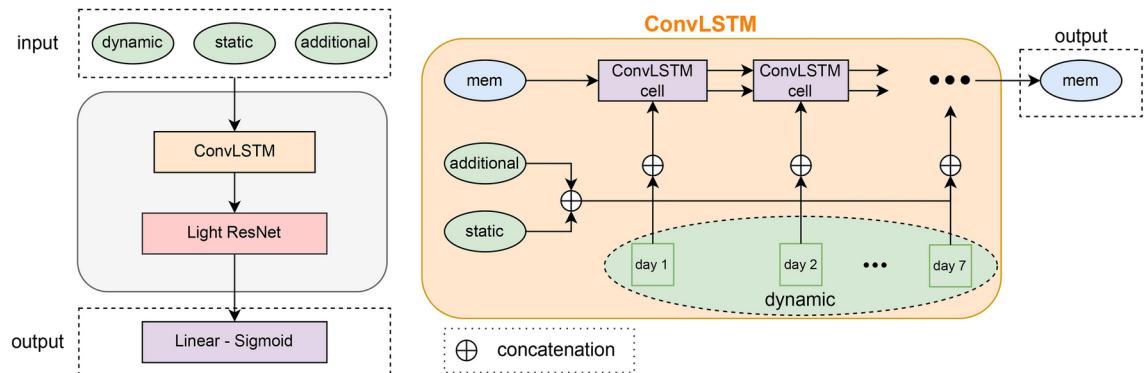


Fig. 11. ConvLSTM network. Left side is general network structure, rights side is ConvLSTM encoder.

Layer	Output shape	Parameters
Linear(106, 32) + BN + ReLU	(BS, 32)	3426
Linear(32, 32) + BN + ReLU	(BS, 32)	1058
Linear(32, 32) + BN + ReLU	(BS, 32)	1058
Linear(32, 16) + BN + ReLU	(BS, 16)	530
Linear(16, 4)	(BS, 4)	68
Linear(4, 16) + BN + ReLU	(BS, 16)	82
Linear(16, 32) + BN + ReLU	(BS, 32)	546
Linear(32, 32) + BN + ReLU	(BS, 32)	1058
Linear(32, 32) + BN + ReLU	(BS, 32)	1058
Linear(32, 106)	(BS, 106)	3498

Table 4. An autoencoder architecture implemented for anomaly detection. BN is a trainable batch normalization module, BS - minibatch size.

- Changing the dimensionality of the input tensor (necessary due to the fact that we use complex structured features for training);
- Adding a lightweight convolutional network in addition to ConvLSTM to bring the dimensionality of the output tensor to the desired format.

3. Anomaly detection algorithms

MLP Autoencoder (AE). We implemented a Multi-Layer Perceptron (MLP) Autoencoder for 5 days wildfire occurrence prediction. The AE consists of an encoder, which compresses input features into a latent space, and a decoder, which reconstructs the input. The reconstruction error serves as an anomaly score, where higher errors indicate deviations potentially corresponding to fire events. The detailed AE architecture is presented in Table 4.

Metrics

The following metrics are usually used to evaluate the quality of the classification models:

1. Precision = $\frac{TP}{TP+FP}$
2. Recall = $\frac{TP}{TP+FN}$
3. F1-Score = $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
4. Specificity = $\frac{TN}{TN+FP}$

Considering the presence of a natural class imbalance in the task, to obtain a more objective evaluation, we calculate the F1-score not only for the positive class but also for the negative class, and then look at the average of the two F1-scores. Also, an important indicator of the model's predictive capabilities is the visual analysis of the map with the forecast for an entire region over a specific period of time.

To assess the quality of models, we also proposed a custom metric called $F1_{balanced}$ designed specifically for assessment of wildfire prediction. This metric is based on calculations of the F1 measure on random balanced subsamples of the test dataset. For the area of interest for each day, a set of samples is formed, consisting of all examples with fires and the same number of randomly selected examples without fires. In this way, a balanced set of samples is formed for the entire time period $F1'_{balanced}$. $F1'_{balanced}$ is calculated N times with different random seed and then averaged:

$$F1_{balanced} = \frac{1}{N} \sum_1^N F1'_{balanced} \quad (4)$$

Additionally, we use the F1 metric in the same way as in³⁶:

$$F_{\beta}^+ = \frac{(1 + \beta^2) \text{TP}}{(1 + \beta^2) \text{TP} + \beta^2 \text{FN} + \text{FP}}, \quad F_{\beta}^- = \frac{(1 + \beta^{-2}) \text{TN}}{(1 + \beta^{-2}) \text{TN} + \beta^{-2} \text{FP} + \text{FN}}, \quad F_{\beta} = \frac{F_{\beta}^- + F_{\beta}^+}{2} \quad (5)$$

The F_{β} metric are designed to evaluate the performance of a model while allowing for different weightings of true positives (TP), false negatives (FN), true negatives (TN), and false positives (FP), based on the value of β . As β increases, the emphasis shifts toward the contribution of TP and FN, making the metric more focused on the model's ability to accurately predict positive class.

In fire occurrence prediction task, where class imbalance is common, the F_{β} metric helps to account for this imbalance by placing more emphasis on correctly predicting fire events, ensuring the model's effectiveness in identifying rare but critical occurrences.

Test data sampling approach

A balanced test sample is used to evaluate the model where the number of examples with fire is equal to the number of examples without fire. Here we describe an algorithm for obtaining a balanced test sample for a single prediction date:

1. In the first step, a table of target fires is generated. Each fire is characterised uniquely by its spatial location (id of the grid cell), and its time interval (the date of the start and end of burning) (see section “[Fire points preprocessing](#)”). To obtain target fires, for each forecast cell, all fires whose burning interval has an overlap with the 5 day forecast horizon are selected.
2. Next step is selection of representative cells. A cell is considered representative if it is located in areas potentially prone to fires. For selecting these cells, information on fires from 2012 to 2022 is used: a region with a radius of 100 km is constructed around each fire, and the resulting area is the union of all these regions. The grid containing the representative cells is then saved.
3. The last step involves balancing the samples. All cells corresponding to fires are taken (let the number of cells with fires be k), then k representative cells without fires are selected using the grid obtained in the previous step.

In the case of forecasts for multiple dates, the second stage of data preparation is conducted for each forecast date and the resulting datasets are combined. In this way, we obtain a balanced dataset for testing the effectiveness of the model that does not contain overtly 'non-burning' examples.

Implementation details

Dataset module architecture

For the convenience of training various models, a universal dataset module has been implemented, which allows the use of preprocessed data. The dataset module is a program that allows us to load data in a special format from the storage based on specified geographic boundaries, date, and some other parameters. Our implementation uses the output of 3 tensors: (1) static features (including data on topography of the area, population for the current year, land cover type and vegetative indices for the last 2 weeks); (2) Dynamic features (daily weather data); and (3) optional additional features (day of year, absolute coordinates of the center of the uploaded patch). For the training mode, it is also feasible to include a label on the output tensors that indicates whether there will be a fire in the area under consideration in the next N days. This label distinguishes between the presence of a fire (binary class '1') and the absence of a fire (binary class '0'). The main settings of the dataset module are described below:

- *sample_raster_size* is the parameter that specifies the spatial dimensions of the output tensors in pixels;
- *day_seq_length* is the parameter that specifies length of the series (in days) of historical weather data;
- *fire_interval* is the parameter defining the range of days to search for fires in the area under consideration;
- *LC_mode* is the field for selecting the Land Cover Type loading mode. There are two modes are available: (1) One-Hot Encoding - loading Land Cover Type features as a binary tensor of size $17 \times H \times W$, where 17 is the number of Land Cover classes, and $H \times W$ are the spatial dimensions of the tensor, specified by *sample_raster_size*, and (2) Label Encoding - loading as a tensor of size $H \times W$, where each pixel value is given by the Land Cover Class;
- *sample_with_date* is the flag that adds the day of the year as an additional feature (additional output tensor);
- *sample_with_coords* is the flag that adds the absolute coordinates of patch center as an additional feature (additional output tensor);
- *augments* specifies the set of data augmentations applied to the loaded data. The approach we implemented uses 90-180-270 degree rotations and mirror reflections.

The dataset module was used to train ML models, both to generate minibatches for optimizing neural networks using the gradient descent method, and to generate a matrix of samples for training ML methods based on decision trees.

Training details

It is worth noting that all models are trained on the same dataset, but the input format of CNN models differs from the input format of models based on decision trees. CNN models require as input a tensor, one of the slices of which is a raster image of some feature, on the other hand classical ML-based models require a set of features stretched into one vector as input. To take this factor into account, when training some models, median and maximum values of a particular feature were used instead of the full raster (for Land Cover Type, the mode of values was used). For each sample, the values obtained in this way were collected into one vector and used for training. This feature representation method was used to train RF and XGBoost classifiers, as well as to train an anomaly search model based on Auto Encoder.

Random forest (RF). As mentioned above, we used the RF implementation from the open-source scikit-learn package. To obtain optimal hyperparameters of the method, we used the Optuna framework⁷³. Using Optuna, the following near-optimal parameters were obtained: the number of decision trees (*n_estimators*) equal to 223, maximum depth of each decision tree (*max_depth*) equal to 11, the minimum number of samples required to split an internal node (*min_samples_leaf*) equal to 1, and the minimum number of samples required to be at a leaf node (*min_samples_split*) equal to 3.

XGBoost (XG). In the process of training this method based on decision trees, we used the open source implementation of gradient boosting XGBoost and also the Optuna framework for selecting hyperparameters. As a result of searching for the best hyperparameters using Optuna, we obtained the following values: the number of decision trees (*n_estimators*) equal to 100, maximum depth of a tree (*max_depth*) equal to 6 and learning rate equal to 0.1.

Auto encoder (AE). To train the anomaly search model, AE was implemented, operating in the mode of encoding and decoding the feature vector. The implemented AE consists of sequential fully connected layers with ReLU nonlinearity. The AE depth and the number of parameters in each layer were selected as a result of a grid search. The final AE architecture is presented in Table 4. The following strategy is presented for AE training. Samples without fire were selected as non-anomaly examples, and samples with fire were selected as anomalous examples. At the first stage of training, AE was trained to compress input features into latent space and then reconstruct them. This part of training was carried out using minibatch gradient descent with AdamW optimizer and MSE loss function. The second part of setting up the model was to find the optimal error threshold for feature reconstruction to search for anomalies. For this stage, feature reconstruction errors (MSE) were calculated for the entire validation dataset, both for examples with fires and for examples without fires. The best threshold was selected based on the F1-score criterion for separating anomalous and non-anomalous models.

CNN models. A similar training pipeline was used to train the adapted ConvLSTM, AMLP, and trimmed RegNetX. To train neural network models, we used the resources of the ZHORES supercomputer⁷⁴, including the NVIDIA A100 GPU. To train each of the models, the AdamW optimizer was used with a learning rate equal to 10^{-4} , a weight decay equal to 10^{-5} , β_1 equal to 0.9, and β_2 equal to 0.999. During training, we used the binary cross-entropy loss function (BCE). Each neural network was trained until a configuration was found that reached the local minimum of the loss function on the validation set. Most often, about 20 training epochs were sufficient for this.

In Table 5, we analyze the key characteristics related to the computational complexity of previously described convolutional neural networks – ConvLSTM, AMLP and RegNetX. Key metrics, such as the total number of parameters, FLOPs, and the training time for 20 epochs on the NVIDIA A100 GPU averaged across regions, illustrate their computational demands and efficiency. It is noteworthy that despite the substantial differences in FLOPs among the models, the variations in training time remain surprisingly narrow. This pattern suggests that much of the training time is spent on batch preparation, underscoring the crucial role of data management strategies in enhancing computational efficiency alongside architectural considerations.

The comparison table is limited to deep neural networks due to their significant computational complexity and resource intensity compared to other models used in the experiments.

Results

We conducted a comparative analysis of various ML and DL models for predicting the probability of fire occurrence within a 5-day window. To thoroughly evaluate the models, we performed two types of calculations: (1) an evaluation of forecasts across the entire region over a period of 1 to 2 months (depending on the region's size), and (2) an evaluation on a balanced sample over the same time frame. The first approach allows us to assess model performance under conditions of significant class imbalance, which is typical of the problem and aligns with real-world application scenarios. In the second approach, the sample is balanced (as detailed in section [Test data sampling approach](#)), providing a more equitable assessment of the model's error contribution across each class.

Model	Parameters	FLOPs (B)	Training time (h)
AMLP	283,777	0,005	11,91
ConvLSTM	366,929	7,674	21,83
RegNetX	377,033	0,108	17,55

Table 5. Computational characteristics of CNN models.

The results of the regional forecasts are summarized in the following tables for Amur, Irkutsk, Rostov, and Sverdlovsk Oblasts (Table 6). In addition to the well-known metrics such as TP , FN , TN , FP , these tables also include F_β metrics ($\beta = 1, 5, 20, 100$).

It is evident that no single approach consistently yields the best results across all regions. While different models may excel in certain areas, there is no universally superior model when considering the overall performance metrics. Specifically, DL methods tend to perform better at predicting non-fire events, resulting in fewer false positives and higher true negative rates, as reflected in metrics such as F_1 , TN , and FP . On the other hand, ML methods generally show superior performance in accurately predicting fire occurrences, as indicated by higher values in metrics like $F_{20,100}$, TP , and FN . These patterns suggest that while DL models may be more cautious, reducing false alarms, ML models are more effective in identifying actual fire events, albeit with a different trade-off in false positives and negatives.

Based on the results of the region-wide forecast evaluations, it can be concluded that each region requires a tailored approach when selecting a fire probability forecasting model. The optimal model may vary depending on the unique characteristics of the region and the specific requirements for managing Type I errors (false positives) and Type II errors (false negatives).

To gain a deeper understanding of the reasons behind these varied results, we conducted a series of experiments focused on investigating the underlying factors. These analyses are detailed in the section ‘Discussion’, where we explore how the distinct importance of features and the variability in weather data distributions across regions may influence model performance.

Table 7 summarizes the results of model evaluation on balanced samples, where the number of fire occurrences is equal to the number of non-fire instances.

The evaluation results on these balanced samples are more uniform, with a clear advantage for classical ML approaches, particularly RF and XGBoost. The superior performance of these models can likely be attributed to their robustness to noise and inaccuracies in weather data, as well as their effective aggregation of data across varying resolutions. In addition, as noted above, classical models exhibit stronger predictive performance for fire events, a strength that is highlighted in balanced datasets where equal representation of fire and non-fire cases allows these models to fully exploit their predictive capabilities.

Model	F_1	F_5	F_{20}	F_{100}	TP	FN	TN	FP
Amur Oblast								
AMLP	0.4453	0.8176	0.9714	0.9861	1432	37	28076	18195
ConvLSTM	0.4541	0.8167	0.9621	0.9758	1403	66	28960	17311
RF	0.4419	0.8174	0.9742	0.9893	1441	28	27743	18528
XGBoost	0.4429	0.8168	0.9722	0.9872	1435	34	27854	18417
Auto encoder	0.3445	0.6995	0.8576	0.8743	1126	343	20447	25824
RegNetX	0.4952	0.8319	0.9500	0.9604	1359	110	32606	13665
Irkutsk Oblast								
AMLP	0.3855	0.7252	0.9023	0.9220	1202	208	30858	27402
ConvLSTM	0.4173	0.7215	0.8659	0.8810	1089	321	35122	23138
RF	0.4172	0.7253	0.8725	0.8880	1108	302	35017	23243
XGBoost	0.4581	0.7168	0.8206	0.8305	949	461	40834	17426
Auto encoder	0.3411	0.6724	0.8449	0.8651	1051	359	26395	31865
RegNetX	0.4114	0.7297	0.8856	0.9022	1147	263	34139	24121
Rostov Oblast								
AMLP	0.3342	0.6077	0.8102	0.8437	213	92	11625	13141
ConvLSTM	0.4065	0.6377	0.7982	0.8202	198	107	15632	9134
RF	0.4216	0.6625	0.8314	0.8537	218	87	16374	8392
XGBoost	0.4174	0.6731	0.8600	0.8853	237	68	16004	8762
Auto encoder	0.3951	0.5916	0.7119	0.7285	143	162	15319	9447
RegNetX	0.4434	0.6589	0.7945	0.8109	192	113	17835	6931
Sverdlovsk Oblast								
AMLP	0.5005	0.6759	0.7944	0.8079	192	118	49154	5976
ConvLSTM	0.4729	0.6812	0.8883	0.9190	261	49	44909	10221
RF	0.4669	0.6998	0.9557	0.9965	309	1	43581	11549
XGBoost	0.4955	0.7455	0.9691	0.9987	310	0	46990	8140
Auto encoder	0.4706	0.5527	0.6024	0.6085	69	241	47535	7595
RegNetX	0.4648	0.6671	0.8765	0.9092	255	55	43920	11210

Table 6. Numerical comparison of the quality of the studied approaches across all regions.

Model	Amur			Irkutsk			Rostov			Sverdlovsk		
	$F_1(1)$	$F_1(0)$	$F_1\text{avg}$	$F_1(1)$	$F_1(0)$	$F_1\text{avg}$	$F_1(1)$	$F_1(0)$	$F_1\text{avg}$	$F_1(1)$	$F_1(0)$	$F_1\text{avg}$
AMLP	0.8224	0.7479	0.7852	0.7676	0.6657	0.7166	0.579	0.4706	0.5248	0.7135	0.7732	0.7434
ConvLSTM	0.8182	0.7507	0.7845	0.7417	0.6960	0.7188	0.655	0.6288	0.6419	0.8113	0.7909	0.8011
RF	0.8228	0.7412	0.7820	0.7505	0.6894	0.7199	0.7051	0.671	0.6881	0.8573	0.8017	0.8295
XGBoost	0.8253	0.7479	0.7866	0.7106	0.7136	0.7121	0.7269	0.6693	0.6981	0.8876	0.855	0.8713
Auto Encoder	0.6999	0.584	0.6420	0.6933	0.5586	0.6259	0.4329	0.5636	0.4983	0.3318	0.6636	0.4977
RegNetX	0.8451	0.8063	0.8257	0.7513	0.6875	0.7194	0.6303	0.6522	0.6413	0.8046	0.7909	0.7978

Table 7. Comparison of the quality of the studied approaches on a balanced sample. $F_1(1)$ - $F1_{balanced}$ calculated for class of fire, $F_1(0)$ - $F1_{balanced}$ calculated for inverted target class (no fire), $F_1\text{avg}$ - average of $F_1(1)$ and $F_1(0)$.

Visualization of forecasts

To further analyze the predictive performance of our models, we present visual comparisons of wildfire forecasts for two regions: Amur Oblast (Fig. 12) and Sverdlovsk Oblast (Fig. 13). Each figure illustrates the probability of fire occurrence predicted by different models for a specific day within the 5-day forecasting window. The forecasts are shown as heatmaps, where color intensity represents the predicted probability of wildfire occurrence, and black crosses indicate actual fire events observed on the forecasted dates. For each region, the subplots correspond to different models, allowing a direct visual comparison of their predictions. The title of the subplot representing the best-performing model is explicitly marked to help clarify. This visualization helps highlight spatial patterns in predictions, as well as differences in model sensitivity to wildfire-prone areas.

In the Amur Oblast, all models except Auto Encoder demonstrate high recall, meaning they successfully identified most wildfire occurrences. Among them, the RegNetX model stands out due to its superior precision, as it produces the fewest false alarms compared to other models. This aligns with the numerical results presented in Table 6, where it achieves the best true negative (TN) and false positive (FP) scores for this region. In the Sverdlovsk region, Random Forest, ConvLSTM, and Attention MLP produce numerous false fire predictions, while RegNetX and Gradient Boosting have comparable false positive rates. However, RegNetX misses some isolated wildfires near the 60th latitude, whereas Gradient Boosting accurately predicts all fire occurrences with minimal false positives. This is consistent with Table 6, confirming Gradient Boosting as the most reliable model for this region. For a complete view, visualizations of the top performing models for each region are presented in the Supplementary Information.

The dataset and code example is available through the link <https://github.com/LanaLana/Wildfire-Forecasting>.

Discussion

Significance of features by region

In this experiment, we focused on analyzing the importance of features for the XGBoost model, as it consistently delivers strong results and offers easy interpretability. The primary goal was to assess how models trained on data from different regions differ in terms of the features they consider important. Specifically, we aimed to understand which meteorological and environmental factors are most significant for predicting fires in various climatic zones.

XGBoost evaluates feature importance using three main approaches: Gain, Cover, and Weight, each providing a different perspective on the contribution of features to the model's decision-making process.

- *Gain* measures the average reduction in the loss function achieved by using a feature for splits across all trees. A higher Gain indicates that the feature contributes more to minimizing prediction errors.
- *Cover* represents the number of training samples affected by a given feature across all decision trees. It indicates how widely a feature is utilized for making splits, regardless of the quality of splits.
- *Weight* counts the number of times a feature is used to split the data across all trees. However, it does not account for the split's effectiveness. A high Weight suggests that the feature is commonly relied upon by the model.

Among these methods, Gain-based importance is the most useful for assessing the true impact of each feature, as it directly measures how much each feature improves the model's decision-making. This metric is robust to feature cardinality and avoids the biases inherent in Weight and Cover. Therefore, in this study, we primarily rely on Gain to analyze the relative importance of features across different regions. The feature importance plots for each region are presented in Fig. 14.

Regions of Eastern Siberia - namely, Amur and Irkutsk, are characterized by extensive forest coverage and complex terrain. In both regions, *Nesterov index* is a key indicator, capturing the cumulative impact of heat and drought on wildfire risk. In Amur, the significance of *Elevation* is linked to its mountainous terrain, which shapes microclimate and fuel availability. In Irkutsk, the importance of *PET* and *Total precipitation* highlights the role of soil moisture in reducing wildfire risk in plateau forests. *NDVI* feature, which evaluates vegetation density and health, is relevant in both regions. As we demonstrated the importance of vegetation, particularly the properties of forest cover, additional and more detailed characteristics may provide valuable information for

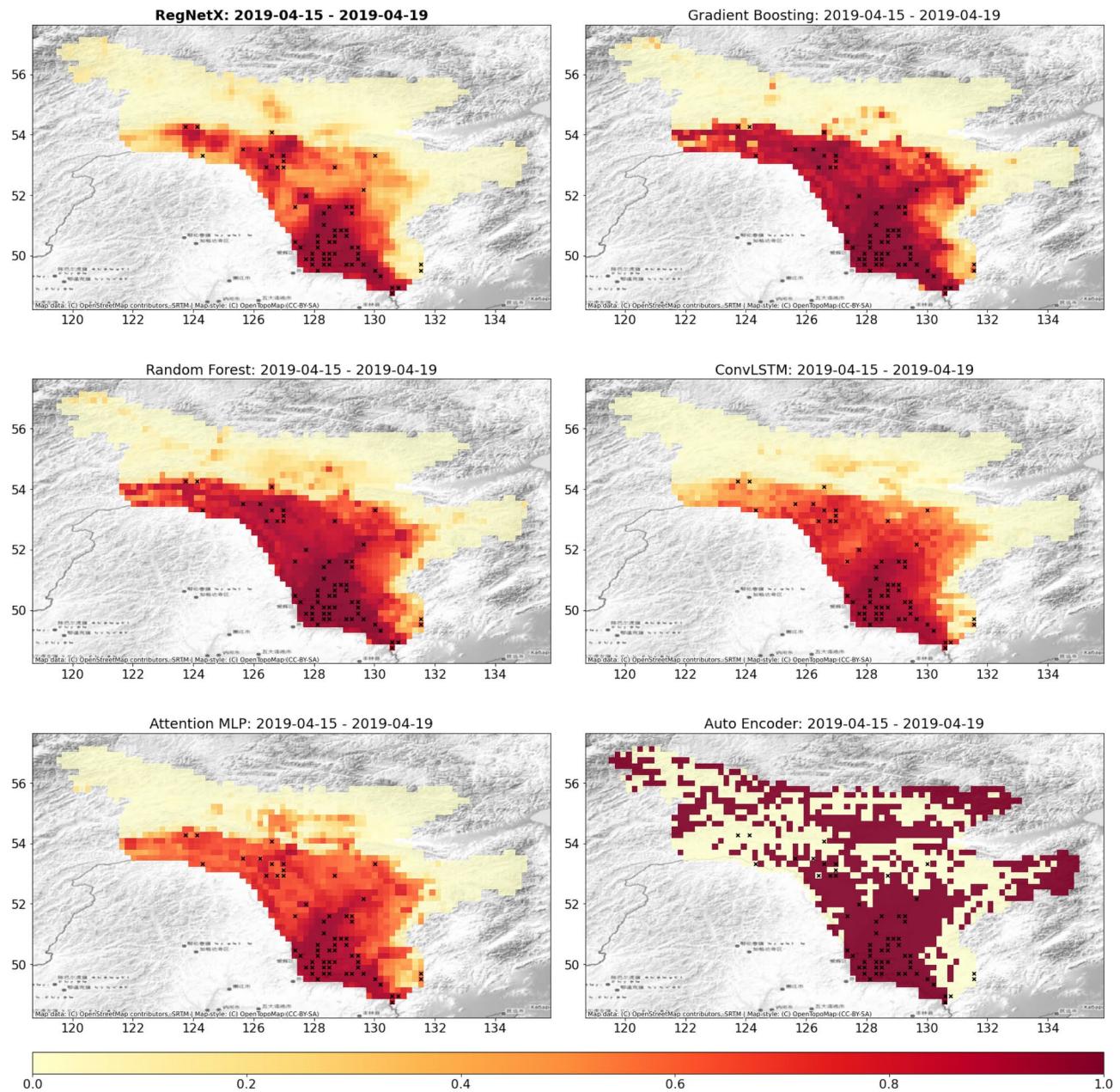


Fig. 12. Comparison of models in the Amur Oblast. Each subplot corresponds to a 5-day forecast by a specific model, the red color scale corresponds to the predicted fire probability, black crosses are real fires. The best performing model is indicated by a bold title. The figures were created by the authors using Matplotlib library version 3.5.3 (<https://matplotlib.org/>) on Python 3.10 version, OpenTopoMap composite derived from OpenStreetMap (<https://opentopomap.org/about>) is chosen for visualization.

future studies. Rostov Oblast, characterized by its arid climate, flat terrain and high population density, highlights the importance of *ET* as an indicator of drought. Human influence is captured through *Population* feature, while *Land Cover* defines the diversity of fuel sources, including agricultural land and pastures. In contrast, *Elevation*, *NDVI*, and *Nesterov index* are less significant due to the flat landscape and limited forest coverage. Sverdlovsk Oblast, with its temperate continental climate and extensive forest coverage, shares similarities with eastern regions in the importance of *Nesterov index*. *Land Cover* is essential for evaluating wildfire risk across diverse landscapes. Meanwhile, features like *Total precipitation*, *ET* and *Elevation* are less relevant, reflecting the region's stable precipitation patterns and limited slope effects in predominantly forested areas.

The analysis demonstrated that feature importance is closely related to the environmental factors of each region, suggesting that different climatological zones require specific approaches to fire prediction.

Building on the insights gained from the feature importance analysis with XGBoost, we observed that different regions prioritize distinct meteorological and environmental factors in fire prediction. This variability prompted us to further investigate whether the differences in these influential features are reflected in the underlying

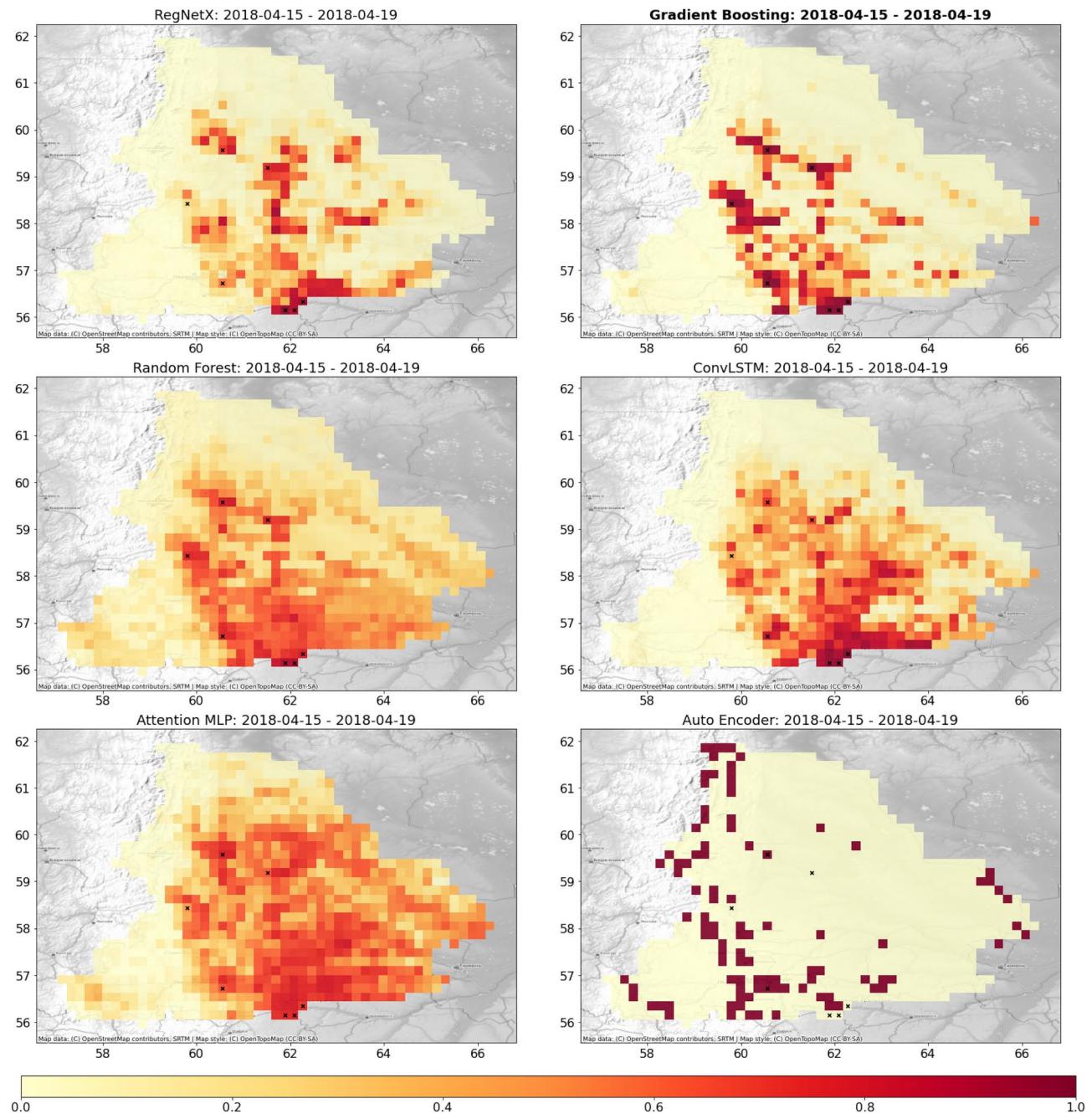


Fig. 13. Comparison of models in the Sverdlovsk Oblast. Each subplot corresponds to a 5-day forecast by a specific model, the red color scale corresponds to the predicted fire probability, black crosses are real fires. The best performing model is indicated by a bold title. The figures were created by the authors using Matplotlib library version 3.5.3 (<https://matplotlib.org/>) on Python 3.10 version, OpenTopoMap composite derived from OpenStreetMap (<https://opentopomap.org/about>) is chosen for visualization.

distributions of meteorological data across regions and time periods relative to fire events. To explore this, we conducted a detailed comparison of the distributions of weather features, analyzing how these distributions vary as a function of temporal distance from the day of a fire. This approach allows us to assess whether specific patterns in meteorological data can consistently predict fire occurrence across different regions.

Distribution of meteorological data

A week's supply of historical data is used to predict the probability of a fire occurring the following days. Considering the frequency of change in the data used, it can be argued that that meteorological characteristics will be the primary source of information regarding the likelihood of a fire on the subsequent days for any given forecast cell.

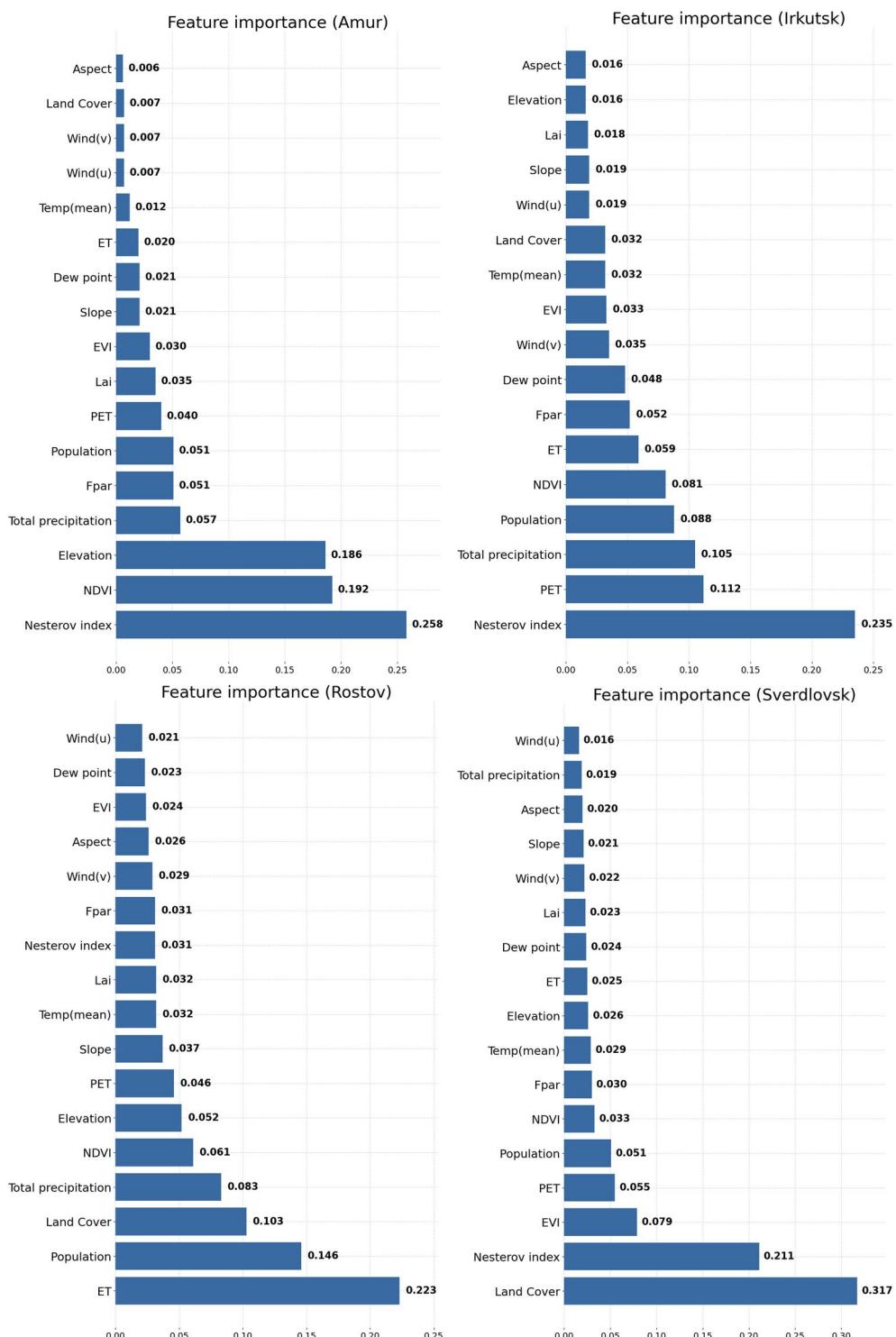


Fig. 14. XGBoost feature importance for different regions.

The primary goal of this experiment is to determine whether there is a significant difference in the distributions of weather features as a function of temporal distance from the day of the fire. The results of this experiment aim to reveal the predictive potential of meteorological data in forecasting fire occurrences.

For this experiment, data for 2018, 2019 was employed. There are six meteorological features: u and v components of wind, air temperature, dew point temperature, total amount of precipitation and Nesterov index. For each fire event, meteorological attribute values are collected over a seven-day period, with varying offsets from the fire day: $k = 0, 1, 3, 5$.

Example We know that the fire started on 13 May, then an offset of k days corresponds to data taken from 13 – 7 – k May to 13 – 1 – k May.

This approach allows the collection of examples in four groups corresponding to the four specified offsets: *fire* ($k = 0$), *no_fire_1* ($k = 1$), *no_fire_3* ($k = 3$), *no_fire_5* ($k = 5$). The distributions of meteorological features from these groups are then compared using both visual and statistical analysis methods. It is hypothesized that as k increases, the difference between the distributions of meteorological data for the 0 and k day gaps will also increase, indicating a stronger predictive signal closer to the day of the fire.

To account for potential variations due to different climate zones, the data were stratified by regions. Number of fire samples for regions: 1541 fires for Amur Region, 803 fires for Irkutsk Region, 337 for Rostov Region, 382 for Orenburg Region. For visual analysis, comparison graphs of distributions and cumulative distribution functions (CDFs) were plotted, as shown in Fig. 15. Here we present figures from one region - the Amur Region - while the key findings from this analysis are discussed in the text. Similar patterns were observed in other regions, with variations in specific features (See Supplementary Figs. 1, 2, 3).

Statistical tests were also conducted to assess the hypothesis that the samples of meteorological attributes from different groups follow the same distribution (i.e. that the meteorological attributes of the two groups are equally distributed). The nonparametric Kolmogorov-Smirnov test was selected as the statistical method for this purpose. This criterion not only allows the hypothesis of distributional equivalence to be tested, but also provides insight into the relationships between distributions.

Note Kolmogorov-Smirnov criterion. The Smirnov uniformity criterion is used to test the hypothesis that two independent samples belong to the same distribution law, that is, that two empirical distributions correspond to the same law. Null hypothesis H_0 : the two studied samples follow the same distribution. In addition to the null hypothesis of equality of distributions ($F(x) = G(x)$), we are interested in knowing the relationship between the distributions, which may indicate the separating potential of the attribute, therefore, in addition to the main hypothesis, less ($F(x) \geq G(x)$) and greater ($F(x) \leq G(x)$) hypothesis were tested.

The results of pairwise comparisons of the distributions of meteorological attributes for different groups (fire, no fire) are presented in the Table 8. Bold text indicates the attributes for which the relation (more, less) between the distributions is obtained.

Based on the analysis, the following conclusions can be drawn regarding meteorological attributes. The Kolmogorov-Smirnov test revealed significant differences in the distributions across all pairs of weather feature groups, suggesting a potential ability to distinguish between these groups. However, this result alone does not imply the ability to accurately classify examples belonging to different groups. Additionally, the comparison graphs of distributions and CDFs show that for at least half of the meteorological features, the distributions appear visually similar. The significant results from the Kolmogorov-Smirnov test might be influenced by factors such as sample size or sensitivity to minor differences, rather than reflecting genuine discriminative power.

In the case when we know the ratio between the distributions of groups (statistically greater or less) we can, with a certain probability, classify a new example based on this feature. Therefore, features that demonstrate statistically significant relationships between group distributions are more informative for predicting the day of fire occurrence. In all regions, at least one feature establishing a relationship between the distributions of most pairs of groups was obtained. Notably, the Nesterov fire hazard index most frequently exhibited these relationships, highlighting its importance in fire prediction. This result is logical, given the theoretical foundation of using the fire hazard index for forecasting fires. The next most significant feature is the amount of precipitation, which aligns well with intuitive expectations regarding fire risk. The findings related to the Nesterov index are particularly consistent with visual observations from the comparison of distributions and CDFs, further reinforcing its significance in the analysis.

Statistical tests, along with visual analysis, provided evidence that certain features can differentiate between all pairs of groups. However, neither approach showed an increase in separability as the temporal offset from the fire day increased. Consequently, the ability to accurately classify the exact day of a fire based solely on weather data remains uncertain and cannot be guaranteed.

Despite the conclusions drawn above, it is a well-known fact that neural network methods trained on large, high-quality datasets achieve high classification accuracy. In our future research, our aim is to continue our analysis by utilizing feature importance methods such as SHAP⁷⁵ and LIME⁷⁶. We believe that understanding the significance of input features in the task at hand can provide valuable insights into the direction of further development.

Conclusion

In this study, we address the critical task of predicting wildfire occurrences using remote sensing data. The primary challenges in developing AI-based solutions for this task stem from the heterogeneity of existing environmental measurements that can influence fire occurrence, as well as the lack of a unified pipeline for data acquisition and processing. To tackle these challenges, we investigated several freely available data sources for meteorological, vegetation, and anthropogenic measurements, and proposed a methodology for developing ML solutions.

The dataset we compiled includes over 17,000 verified wildfire events across four large regions of the northern hemisphere with different topographic and climatic conditions over a span of ten years. Using this dataset, we explored the correlation between various spatial environmental features and the probability of fire occurrence under natural conditions. Experiments showed that both the shape of the distributions of the weather variables considered and the dynamics of their changes can differ significantly. Our findings indicate that model

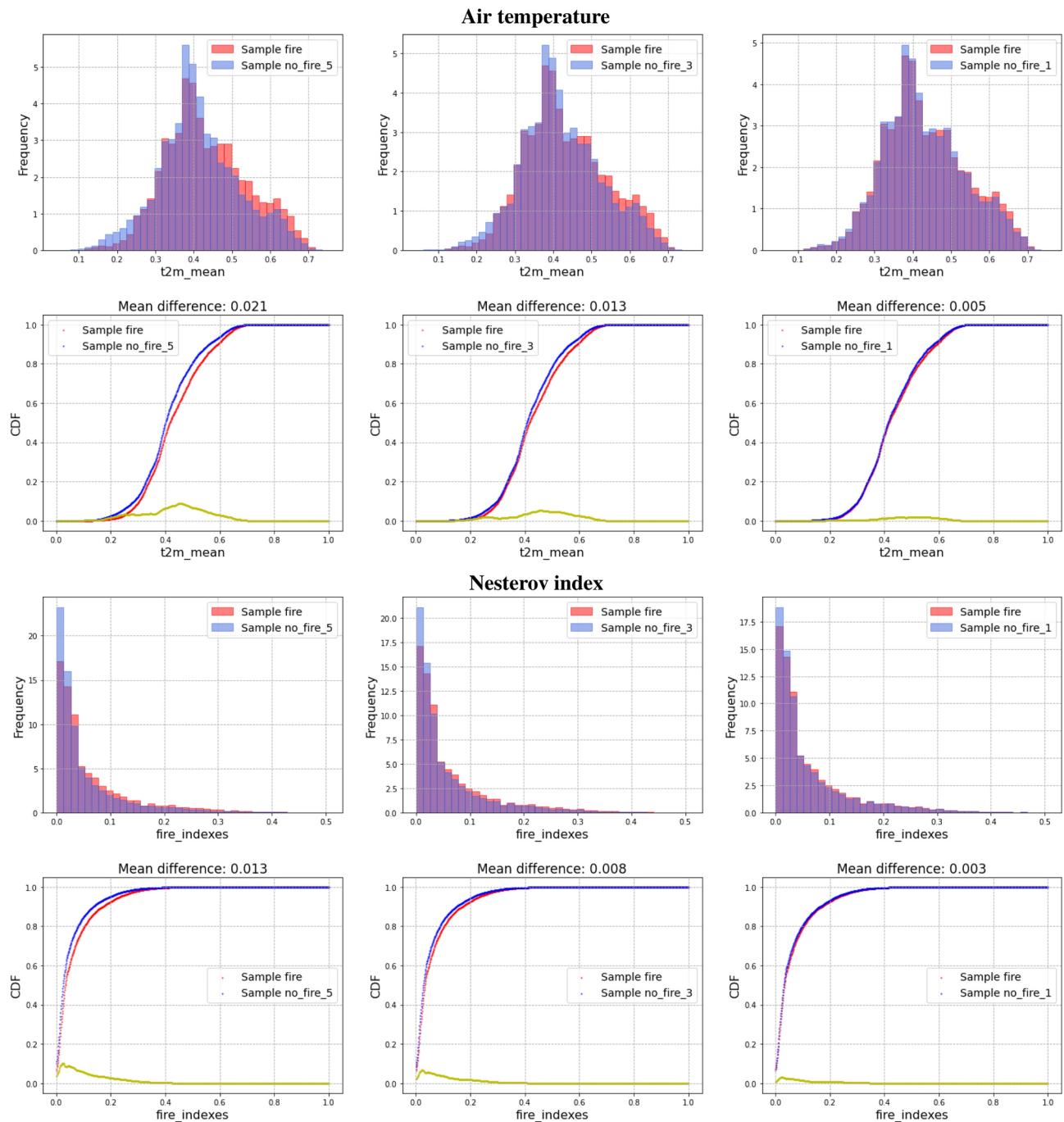


Fig. 15. Distributions of meteorological features of different groups for Amur Region. The yellow graph represents the absolute difference in CDF of the compared groups. *Mean difference* – mean value of the yellow graph.

performance is significantly influenced by feature distribution and environmental conditions, suggesting that it is preferable to select an individual model for each region.

Additionally, we addressed the crucial issue of evaluating ML models in the context of forecasting rare events such as wildfires. We discussed several metrics to provide a deeper understanding of model performance and to represent results in terms of the spatial and temporal distribution of fire events throughout the year. Overall, the proposed methodology encompasses and analyzes key aspects of wildfire emergency system development and validation. It demonstrates significant potential for future expansion to other regions with varying environmental conditions.

	<i>u10_mean</i>	<i>v10_mean</i>	<i>d2m_mean</i>	<i>t2m_mean</i>	<i>tp_sum</i>	<i>fire_index</i>
Amur region						
no_fire_5 VS fire	\neq	\neq	\neq	<	\neq	<
no_fire_3 VS fire	\neq	\neq	\neq	<	>	<
no_fire_1 VS fire	<	\neq	\neq	<	>	<
Irkutsk region						
no_fire_5 VS fire	\neq	\neq	\neq	<	\neq	<
no_fire_3 VS fire	>	\neq	\neq	<	\neq	<
no_fire_1 VS fire	>	<	\neq	<	\neq	<
Rostov region						
no_fire_5 VS fire	>	\neq	\neq	\neq	\neq	<
no_fire_3 VS fire	>	\neq	\neq	\neq	\neq	<
no_fire_1 VS fire	>	>	\neq	\neq	>	<
Sverdlovsk region						
no_fire_5 VS fire	\neq	\neq	\neq	\neq	\neq	\neq
no_fire_3 VS fire	\neq	\neq	\neq	\neq	\neq	<
no_fire_1 VS fire	\neq	\neq	\neq	\neq	>	<

Table 8. Kolmogorov-Smirnov criterion results, $\alpha = 0.05$. Abbreviations $u10/v10$ – u- and v-component of wind, $d2m$ – 2m dewpoint temperature, $t2m$ – 2m air temperature, tp – total precipitation, *fire index* – Nesterov index. Bold indicates attributes where the relation (more, less) between distributions is obtained.

Data Availability

The datasets used and analysed during the current study available from the corresponding author on reasonable request.

Received: 14 September 2024; Accepted: 11 March 2025

Published online: 28 March 2025

References

1. Zhong, S. et al. Machine learning: New ideas and tools in environmental science and engineering. *Environ. Sci. Technol.* **55**, 12741–12754 (2021).
2. Illarionova, S. et al. A survey of computer vision techniques for forest characterization and carbon monitoring tasks. *Remote Sensing* **14**, 5861 (2022).
3. Alkhatib, R., Sahwan, W., Alkhatib, A. & Schütt, B. A brief review of machine learning algorithms in forest fires science. *Appl. Sci.* **13**, 8275 (2023).
4. de Almeida Pereira, G. H., Fusioka, A. M., Nassu, B. T. & Minetto, R. Active fire detection in landsat-8 imagery: A large-scale dataset and a deep-learning study. *ISPRS J. Photogramm. Remote Sensing* **178**, 171–186 (2021).
5. Rauste, Y. et al. Satellite-based forest fire detection for fire control in boreal forests. *Int. J. Remote Sens.* **18**, 2641–2656 (1997).
6. Kinaneva, D., Hristov, G., Raychev, J. & Zahariev, P. Early forest fire detection using drones and artificial intelligence. In *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1060–1065 (IEEE, 2019).
7. Yandouzi, M. et al. Review on forest fires detection and prediction using deep learning and drones. *J. Theor. Appl. Inf. Technol.* **100**, 4565–4576 (2022).
8. Sánchez Sánchez, Y., Martínez-Graña, A., Santos Francés, F. & Mateos Picado, M. Mapping wildfire ignition probability using sentinel 2 and lidar (jerte valley, Cáceres, Spain). *Sensors* **18**, 826 (2018).
9. Seydi, S. T., Hasanlou, M. & Chanussot, J. A quadratic morphological deep neural network fusing radar and optical data for the mapping of burned areas. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sensing* **15**, 4194–4216. <https://doi.org/10.1109/JSTARS.2022.3175452> (2022).
10. Shadrin, D. et al. Wildfire spreading prediction using multimodal data and deep neural network approach. *Sci. Rep.* **14**, 2606 (2024).
11. Cheng, S. et al. Parameter flexible wildfire prediction using machine learning techniques: Forward and inverse modelling. *Remote Sensing* **14**, 3228 (2022).
12. Huot, F. et al. Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–13 (2022).
13. Pang, B. et al. Fire-image-densenet (fidn) for predicting wildfire burnt area using remote sensing data. *Comput. Geosci.* **195**, 105783. <https://doi.org/10.1016/j.cageo.2024.105783> (2025).
14. Oliveira, S. et al. Assessing risk and prioritizing safety interventions in human settlements affected by large wildfires. *Forests* **11**, 859 (2020).
15. Scott, J. H. *Assessing crown fire potential by linking models of surface and crown fire behavior*. 29 (US Department of Agriculture, Forest Service, Rocky Mountain Research Station, 2001).
16. Matthews, S. A comparison of fire danger rating systems for use in forests. *Aust. Meteorol. Oceanogr. J.* **58**, 41 (2009).
17. Taylor, S. W. & Alexander, M. E. Science, technology, and human factors in fire danger rating: The Canadian experience. *Int. J. Wildland Fire* **15**, 121–135 (2006).
18. Hollis, J. J. et al. Introduction to the Australian fire danger rating system. *Int. J. Wildland Fire* **33** (2024).
19. San-Miguel-Ayanz, J., Barbosa, P., Schmuck, G., Libertà, G. & Meyer-Roux, J. *The European forest fire information system (effis)* (Innovative Concepts and Methods in Fire Danger Estimation, In Joint Workshop of Earsel SIG and GOFC/GOLD, 2003).
20. Van Wagner, C. E. et al. *Structure of the Canadian forest fire weather index* Vol. 1333 (Environment Canada, Forestry Service Ottawa, ON, Canada, 1974).

21. Torres, F., Lima, G. S., Martins, S. V. & Valverde, S. R. Analysis of efficiency of fire danger indices in forest fire prediction. *Revista árvore* **41**, e410209 (2017).
22. Abid, F. A survey of machine learning algorithms based forest fires prediction and detection systems. *Fire Technol.* **57**, 559–590 (2021).
23. Safavian, S. R. & Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* **21**, 660–674 (1991).
24. Tong, Q. & Gernay, T. Mapping wildfire ignition probability and predictor sensitivity with ensemble-based machine learning. *Nat. Hazards* **119**, 1551–1582 (2023).
25. Malik, A. et al. Data-driven wildfire risk prediction in northern California. *Atmosphere* **12**, 109 (2021).
26. Qiu, L., Chen, J., Fan, L., Sun, L. & Zheng, C. High-resolution mapping of wildfire drivers in California based on machine learning. *Sci. Total Environ.* **833**, 155155 (2022).
27. Shmuel, A. & Heifetz, E. Global wildfire susceptibility mapping based on machine learning models. *Forests* **13**, 1050 (2022).
28. Ismail, F. N., Sengupta, A., Woodford, B. J. & Licorish, S. A. A comparison of one-class versus two-class machine learning models for wildfire prediction in California. In *Australasian Conference on Data Science and Machine Learning*, 239–253 (Springer, 2023).
29. Khan, S. S. & Madden, M. G. One-class classification: Taxonomy of study and review of techniques. *Knowl. Eng. Rev.* **29**, 345–374 (2014).
30. De Vasconcelos, M. P., Silva, S., Tome, M., Alvim, M. & Pereira, J. C. Spatial prediction of fire ignition probabilities: comparing logistic regression and neural networks. *Photogramm. Eng. Remote. Sens.* **67**, 73–81 (2001).
31. Ntinopoulos, N., Sakellariou, S., Christopoulou, O. & Sfougaris, A. Fusion of remotely-sensed fire-related indices for wildfire prediction through the contribution of artificial intelligence. *Sustainability* **15**, 11527 (2023).
32. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
33. Werbos, P. J. Backpropagation through time: What it does and how to do it. *Proc. IEEE* **78**, 1550–1560 (1990).
34. Sadatrazavi, A., Motlagh, M. S., Noorpoor, A. & Ehsani, A. H. Predicting wildfires occurrences using meteorological parameters. *Int. J. Environ. Res.* **16**, 106 (2022).
35. Jo, H.-W. et al. Modeling historical and future forest fires in south Korea: The flam optimization approach. *Remote Sensing* **15**, 1446 (2023).
36. Bergado, J. R., Persello, C., Reinke, K. & Stein, A. Predicting wildfire burns from big geodata using deep learning. *Saf. Sci.* **140**, 105276 (2021).
37. Shi, X. et al. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* **28**, (2015).
38. Prapas, I. et al. Deep learning methods for daily wildfire danger forecasting. arXiv preprint [arXiv:2111.02736](https://arxiv.org/abs/2111.02736) (2021).
39. Girtsou, S., Apostolakis, A., Giannopoulos, G. & Kontoes, C. A machine learning methodology for next day wildfire prediction. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 8487–8490 (IEEE, 2021).
40. Zeiler, M. D. & Fergus, R. Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part I* **13**, 818–833 (Springer, 2014).
41. Climate: Amur oblast. <https://en.climate-data.org/asia/russian-federation/amur-oblast-891/>. Accessed on 21 August 2024.
42. Average annual resident population. <https://showdata.gks.ru/report/278930/>. Accessed on 21 August 2024.
43. Area of the territory of the Russian Federation by constituent entities of the Russian Federation as of January 1, 2009. https://rosstat.gov.ru/bgd/reg/b09_109/IssWWW.exe/Stg/d01/tbl-01.xls. Accessed on 21 August 2024.
44. Climate: Irkutsk oblast. <https://en.climate-data.org/asia/russian-federation/irkutsk-oblast-791/>. Accessed on 21 August 2024.
45. Climate: Rostov oblast. <https://en.climate-data.org/asia/russian-federation/rostov-oblast-670/>. Accessed on 21 August 2024.
46. Climate: Sverdlovsk oblast. <https://en.climate-data.org/asia/russian-federation/sverdlovsk-oblast-721/>. Accessed on 21 August 2024.
47. Lin, X., Li, Z., Chen, W., Sun, X. & Gao, D. Forest fire prediction based on long-and short-term time-series network. *Forests* **14**, 778 (2023).
48. Nami, M., Jaafari, A., Fallah, M. & Nabiuni, S. Spatial prediction of wildfire probability in the Hyrcanian ecoregion using evidential belief function model and GIS. *Int. J. Environ. Sci. Technol.* **15**, 373–384 (2018).
49. European Space Agency, S. Copernicus global digital elevation model. distributed by opentopography, <https://doi.org/10.5069/G9028PQB> (2021). Accessed on 28 February 2024.
50. Pham, B. T. et al. Performance evaluation of machine learning methods for forest fire modeling and prediction. *Symmetry* **12**, 1022 (2020).
51. Naderpour, M., Rizeei, H. M. & Ramezani, F. Forest fire risk prediction: A spatial deep neural network-based framework. *Remote Sensing* **13**, 2513 (2021).
52. Worldpop hub. population density. <https://hub.worldpop.org/project/categories?id=18>. Accessed on 28 February 2024.
53. The application for extracting and exploring analysis ready samples. <https://appears.earthdatacloud.nasa.gov>. Accessed on 28 February 2024.
54. Modis/terra+aqua land cover type yearly l3 global 500 m sin grid. <https://lpdaac.usgs.gov/products/mcd12q1v061>. Accessed on 28 February 2024.
55. Modis/terra vegetation indices 16-day l3 global 500 m sin grid. <https://lpdaac.usgs.gov/products/mod13a1v061/>. Accessed on 28 February 2024.
56. Qu, Y., Miralles, D. G., Veraverbeke, S., Vereecken, H. & Montzka, C. Wildfire precursors show complementary predictability in different timescales. *Nat. Commun.* **14**, 6829 (2023).
57. Modis/terra net evapotranspiration gap-filled 8-day l4 global 500 m sin grid. <https://lpdaac.usgs.gov/products/mod16a2gv061/>. Accessed on 28 February 2024.
58. Modis/terra leaf area index/fpar 8-day l4 global 500 m sin grid. <https://lpdaac.usgs.gov/products/mod15a2hv061/>. Accessed on 28 February 2024.
59. Qin, H., Wang, C., Zhao, K. & Xi, X. Estimation of the fraction of absorbed photosynthetically active radiation (fpar) in maize canopies using lidar data and hyperspectral imagery. *PLoS ONE* **13**, e0197510 (2018).
60. Bréda, N. J. Ground-based measurements of leaf area index: A review of methods, instruments and current controversies. *J. Exp. Bot.* **54**, 2403–2417 (2003).
61. Muñoz-Sabater, J. Era5-land hourly data from 1950 to present. copernicus climate change service (c3s) climate data store (cds), <https://doi.org/10.24381/cds.e2161bac> (2019). Accessed on 28 February 2024.
62. osgeo.gdal module. <https://gdal.org/api/python/osgeo.gdal.html>. Accessed on 1 March 2024.
63. Buslaev, A. et al. Albumentations: Fast and flexible image augmentations. *Information* **11**, 125 (2020).
64. Belguti, M. & Drágut, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote. Sens.* **114**, 24–31 (2016).
65. Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794 (2016).
66. Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K. & Dollár, P. Designing network design spaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10428–10436 (2020).
67. Li, L. et al. Multistep forecasting of soil moisture using spatiotemporal deep encoder-decoder networks. *J. Hydrometeorol.* **23**, 337–350 (2022).

68. Ahmad, R., Yang, B., Ettlin, G., Berger, A. & Rodriguez-Bocca, P. A machine-learning based convlstm architecture for ndvi forecasting. *Int. Trans. Oper. Res.* **30**, 2025–2048 (2023).
69. Dehghani, A. et al. Comparative evaluation of lstm, cnn, and convlstm for hourly short-term streamflow forecasting using deep learning approaches. *Eco. Inform.* **75**, 102119 (2023).
70. De, S. et al. Griffin: Mixing gated linear recurrences with local attention for efficient language models. arXiv preprint [arXiv:2402.19427](https://arxiv.org/abs/2402.19427) (2024).
71. Hackeling, G. Mastering machine learning with scikit-learn. *Packt Publishing Ltd* (2017).
72. Extreme gradient boosting. python package. https://xgboost.readthedocs.io/en/stable/python/python_api.html. Accessed on 09 September 2024.
73. Optuna. a hyperparameter optimization framework. <https://optuna.readthedocs.io/en/stable/>. Accessed on 09 September 2024.
74. Zacharov, I. et al. “zhores”-petaflops supercomputer for data-driven modeling, machine learning and artificial intelligence installed in skolkovo institute of science and technology. *Open Eng.* **9**, 512–520 (2019).
75. Scott, M. et al. A unified approach to interpreting model predictions. *Adv. Neural. Inf. Process. Syst.* **30**, 4765–4774 (2017).
76. Ribeiro, M. T., Singh, S. & Guestrin, C. “why should i trust you?” explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144 (2016).

Acknowledgements

The work was partially supported by the Russian Science Foundation (Project No. 23-71-01122) and Artificial Intelligence Center of Skoltech.

Author contributions

S.I. conceived the experiment(s), F.G. and M.S. conducted the experiment(s), K.E. and U.T. data curation, S.I., D.S., F.G., M.S., U.T., M.M. and E.B. analysed the results. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.I.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025