FOREST FIRE PREDICTION FRAMEWORKS USING FEDERATED LEARNING AND INTERNET OF THINGS (IOT)

by

Richard Purcell

Submitted in partial fulfillment of the requirements
for the degree of Master of Computer Science

at

Dalhousie University
Halifax, Nova Scotia
August 2023

Table of Contents

List of Tables

List of Figures

Abstract

Forest fires are increasing in severity and frequency due to factors such as climate change. Forest fire ignition prediction is important for safeguarding communities and allocating firefighting resources effectively. Machine learning is a promising tool to enhance forest fire prediction. However, progress in this field is hindered by a scarcity of forest fire ignition datasets for simulation and training. Moreover, existing platforms, such as the Canadian Wildland Fire Information System, underutilize machine learning capabilities for forest fire ignition classification prediction.

To address the limitations in the availability of forest fire ignition datasets and the untapped potential of machine learning within existing platforms, we propose two novel frameworks. The first framework aims to generate comprehensive forest fire ignition prediction datasets for diverse geographic regions. Leveraging R and Python, we instantiate the framework by creating datasets for three fire-prone regions in Canada using existing climate and ignition data. The second framework incorporates federated machine learning and Internet of Things technologies to enable efficient and accurate forest fire ignition classification prediction. By deploying distributed IoT weather stations in a fire-prone region, our federated system harnesses data from multiple sources to enhance prediction capabilities.

To evaluate the effectiveness of the generated datasets, we assess various machine learning models, including Random Forest, Support Vector Machine, and multiple Recurrent Neural Network models. The instantiated datasets yield promising results, achieving an F1-score of 0.85, an Area Under the Receiver Operating Characteristic Curve of 0.95, and a Specificity of 0.89. Additionally, our federated machine learning system achieves comparable prediction accuracy to the centralized system, using an imbalanced dataset with 5,008,365 non-ignitions and 45,411 ignitions. Notably, the federated system achieves superior spatial granularity in classifying forest fire ignitions. These results underscore the potential of our frameworks to enhance forest fire prediction and management, paving the way for improved safety and resource allocation in the face of the increased severity and frequency of forest fires.

Glossary

AB Alberta is a Canadian province, situated between the provinces of Saskatchewan and British Columbia. 23

BC British Columbia is the western most province of Canada. vii, 27, 29

CA The Central Aggregator in our proposed FL system uses simple averaging to combine the weights of WS models during a round. vii, 5, 6, 47

CFFWI The Canadian Forest Fire Weather Index, is an empirical model developed in 1987. This model generated calculations for each weather data collection point, providing insights into various factors related to fire, such as ignition ease, potential fuel consumption, effects of drought on forest fuels, fire spread rate, and overall fuel availability. 9

CS The Central Server is the server were centralized classic machine learning occurs. In our simulation the CS system results are compared to the FL system. 4, 7

CWFIS The Canadian Wildland Fire Information System is a platform developed under the auspices of Natural Resources Canada to assist forest fire management agencies and the public. 4–6, 8, 9, 11, 12, 20, 48

FF Forest fires are an uncontrolled burning of vegetation and trees. They are also referred to as wildfires in some research. 1, 2, 4, 7–9, 12, 15, 19–22, 24, 32, 33, 44–47, 57, 59, 61

FL Federated Learning involves training individual clients on their respective datasets, with only model updates shared and averaged to create a global model. 1, 4–7, 21, 46–51, 53–55, 57, 58, 60, 61

IoT The Internet of Things is a network of interconnected physical devices embedded with sensors, software, and connectivity. ii, v, viii, 1, 4–8, 12, 20, 47–52, 58–61

KNN KNN stands for k-nearest neighbors. It is a machine learning model used for classification and regression. In this model predictions are made based on how similar a new sample is to k number of points in the training data set. 16

LSTM Long Short-Term Memory models are a type of recurrent neural network (RNN). They selectively retain or forget data over sequences, enabling them to capture temporal patterns and dependencies. 5, 17, 47, 54, 55, 58

MB Manitoba is a Canadian province, situated between the provinces of Ontario and Saskatchewan. 23

ML Machine learning is a subsection of artificial intelligence that develops models that enable computers to learn from data, identify patterns, and make predictions. 1, 2, 4, 7, 8, 15, 20, 33, 34, 38, 46, 47, 57–61

NetCDF Network Common Data Form is a set of software libraries made up of data formats that support the creation, access, and sharing of array-oriented scientific data. In our case the form a container for multiple GeoTIFF files. vii, 24, 25, 27, 30

WGCFFDRS The Weather Guide for the Canadian Forest Fire Danger Rating System is a guide on how to interpret and apply weather information in the context of the Canadian forest fire danger rating system. 4, 11, 12

WS Weather Stations within this thesis are those described in [25], automatic stations capable of collecting weather information and reporting it to other devices by radio, satellite, or internet. v, vii, viii, 1, 2, 4–7, 47–52, 54, 58, 60

Acknowledgements

I extend my heartfelt thanks to Dr. Srini Sampalli, Dr. Kshirasagar Naik, Dr. Marzia Zaman, Dr. Chung-Horng Lung, Abdul Mutakabbir, Parveen Kaur, and Fatemeh Tavakoli for their invaluable advice and careful evaluation of my research during our work together on the Predicting Risks of Forest Fires Using Federated Machine Learning Methods project. As well, thanks to the Natural Sciences and Engineering Research Council (NSERC), Canada, for funding the aforementioned project through an Alliance Missions Grant.

Many thanks to my readers, Dr. Manero and Dr. Keselj, for their time and insights in reviewing my research.

The assistance and support of Dr. Darshana Upadhyay, Dr. Jaume Manero, and all members of the Dalhousie MYTech Lab has been very much appreciated. Finally, my family, near and far, thank you for keeping me on track and helping to get this done.

Chapter 1

Introduction

Forest fires (FF) are a natural phenomenon that play a vital role in the regeneration process of healthy forests. However, when FF occur near human communities, they can cause significant physical and economic damage. Therefore, it is crucial to predict the combinations of conditions that can lead to a FF ignition in order to allocate appropriate fire management resources and protect communities. The challenges associated with this task are becoming increasingly complex as forest fires grow in size, frequency, and severity, primarily due to the impacts of climate change. Given the magnitude of this problem, extensive research into all potential solutions is necessary.

To contribute to ongoing research we present two frameworks. The first is for the creation of FF ignition datasets suitable for training machine learning (ML) models. The second is a framework for using Federated Learning (FL), a machine learning method, with Internet of Things (IoT) technologies for classification prediction of FF ignition or non-ignitions in specific geographic areas at a specific times.

1.1 Introduction to a Framework for Creating Forest Fire Ignition Prediction Datasets

We begin by introducing the dataset aspect of this thesis. Publicly available datasets that combine FF ignition data with weather data are sparse and cover limited geographic areas. For instance, Kaggle.com hosts 10 datasets, 7 with no description of the geographic area they cover, and the remainder focused on Montesinho Park in Portugal and the Amazon. Data collection methods are not consistently documented, and sometimes not at all in the FF ignition prediction research we looked at. Hence, a framework is needed that can be used to generate FF ignition prediction datasets. Ideally, this framework is applicable across multiple regions with differing weather features and FF ignition frequencies.

Satellite imagery and averaged numeric features from WS are the principal origins

of data used to train FF ML models. Satellite images are able to capture the majority of the earth's surface but have a limited temporal resolution depending on their orbit. WS are distributed unevenly, clustered close to inhabited areas or sparsely across more remote regions. The temporal and spatial limitations of satellite and weather station collection do not pose a problem for weather prediction, since a well developed and validated model exists [10]. However, in the case of FF ignition prediction which is spatially and temporally specific a general model is not optimal.

We present a framework that provides a general scheme to combine data collected with independent sensing techniques and data derived from reanalysis of historic weather data. The use of reanalysis data is a well accepted practice in extreme weather event and climate research due to its finer temporal and spatial resolutions as noted in Zou et al. [50]. This scheme can fill data holes, providing temporal and spatial data that is more complete. Origins of the historic data include ground based WS, weather balloons, aircraft, ships, and other sources [34]. Data formats from these sources are inconsistent from region to region and time period to time period. The benefit of using reanalysis data is that we are able to generate consistently formatted data for times and geographical areas where data is missing. Ignition data is especially difficult to find in a consistent format. This is due to how the ignition data was collected, the year it was collected in, and the geographical reach of the collecting agency. It is not a trivial task to combine ignition data from one region with that of another, and then with weather data.

Our goal is to simplify the task of creating FF ignition datasets. This is done through a framework and resultant instantiations representing multiple years of ground based weather plus FF ignition data. Our proposed resolution is capable of capturing local variations in temperature, precipitation, humidity, pressure, and FF ignition data over large geographic areas. To benefit from the strengths of ML this data is feature diverse and of a large enough volume that it is representative of a fire prone regions over multiple years. The dataset contributions of this thesis are:

a) A framework for producing high temporal and spatial resolution FF datasets in .csv format with rows representing specific locations and times. Every row contains coordinates, feature values, and a binary column indicating whether an ignition occurred for that location and time. The framework combines weather

reanalysis data (explained in Section 3.2.1) with available ignition data (explained in Section 3.2.2). Provision is made within our framework for other types of data such as proximity to human infrastructure and historical lightning data.

b) Instantiations of our framework resulting in three large datasets suitable for simulating ground-based sensor data:

- A dataset with 19 weather features captured daily over 41 years, 2 spatial features spanning 944,735 km^2, and 5 FF ignition features at a spatial resolution of 0.1° longitude by 0.1° latitude, presented in .csv table format where each row represents a unique day and location.
- Two datasets, each with 19 weather features captured hourly over 41 years, 2 spatial features spanning 254,800 km^2, and 1 FF ignition feature at a spatial resolution of 0.25° longitude by 0.25° latitude, presented in .csv table format where each row represents a unique day and location.

c) Using ML models, we test several fitness aspects of our datasets, including completeness, correctness, and format validity. Modeling the relationship between weather data and ignition data, we show that the dataset is correct (it makes sense spatially and chronologically), complete (models converge), is in a valid format (can be read in a typical data processing scheme), and is acceptable (can be used with multiple models).

d) We use the metrics F1-score, ROC-AUC, Sensitivity, and Specificity to represent several ML models' ability to correctly classify whether an ignition has occurred within a geographic area during a given time interval. The models include bag tree, boost tree, decision tree, logistic regression, multilayer perceptron, nearest neighbor, random forest, and support vector machine (SVM).

e) We demonstrate the usefulness of combining weather reanalysis data with FF ignition data by achieving ROC-AUC and F1-score values above 0.8.

The data framework created has been instantiated for validation, as described in Section 3. Performance of the instantiated datasets has been examined using different

ML models and found to perform with results comparable to past FF prediction research, achieving an F1-score of 0.85, ROC-AUC of 0.95, and Specificity of 0.89.

1.2 Introduction to a Framework using Federated Learning for IoT-based Forest Fire Prediction

We now introduce the FL aspect of this thesis. To contextualize the FL framework we look at the Canadian Wildland Fire Information System (CWFIS), a dynamic platform developed under the auspices of Natural Resources Canada to provide real-time information on FF to fire management agencies and the public [26]. The system continuously enhances its capabilities through the revision of fire weather models, exploration of new methods, and the integration of advanced technologies such as the introduction of MODIS [35] satellite data. However, the current system does not primarily use machine learning (ML) for FF weather analysis and prediction.

The Weather Guide for the Canadian Forest Fire Danger Rating System (WGCFFDRS) emphasizes the need for a larger network of fire WS to ensure reliable fire weather forecasts [25]. While the number of fire WS used by the CWFIS has increased since the 250 stations used in 1995 [26], the emerging impact of climate change, including geographically specific weather phenomena like heat domes, highlights the continued need for an expanded and more geographically diverse network of fire weather collection. The abundance of sensors, like IoT WS, presents an opportunity to address this challenge. By deploying a larger number of WS, it becomes possible to significantly enhance the spatial coverage and resolution of fire weather data collection, enhancing the overall effectiveness of platforms like the CWFIS.

ML is an increasingly explored approach in FF prediction, showing the potential to surpass the predictive power of the empirical models used in the CWFIS [19]. One such approach is FL. FL enables training models on decentralized data without requiring the transfer of data to a CS. The use of FL in FF management has not received extensive research attention, despite its potential to reduce data transfer communication cost within a collection system, and to enhance data sharing through models among fire management agencies. This thesis offers a solution to the question:

Research Question: *How can the integration of FL and IoT technologies*

enhance the capabilities of the CWFIS for efficient FF prediction and management?

There is significant potential to enhance the accuracy of ignition classification prediction models utilized in systems like the CWFIS by harnessing ML techniques, specifically FL, and integrating real-time data from sources such as IoT WSs. Additionally, the creation of a ML model that is shareable at both national and global levels would foster collaboration among various stakeholders engaged in FF prediction.

FL [29] offers several benefits that directly contribute to platforms like the CWFIS. It is characterized by its efficiency, decentralization, and facilitation of collaboration. This thesis introduces a framework that leverages FL to generate predictions using decentralized IoT weather station data, thereby minimizing the need for extensive data transfer and localizing predictions. Simultaneously, it constructs a general model that is shareable with agencies that prefer not to share their data or opt to train a pre-existing model based on local conditions.

Efficiencies provided by FL in the context of FF management manifest in several ways. The framework presented in this thesis suggests the establishment of a network of WS, where each station trains its own ML model, particularly a Long Short-Term Memory LSTM model, and shares the model weights with a central aggregator (CA), Fig. 1.1. The CA averages the weights and sends them back to the stations. This approach eliminates the requirement for extensive data transfer, resulting in reduced bandwidth usage, communication energy consumption, and latency. The output of the framework, a general model for predicting FF ignition enhances data sharing among fire management agencies, even those that wish to silo data. In addition to a framework for FF prediction using FL, this thesis instantiates the framework, demonstrating a possible use case. The FL contributions of this thesis are:

a) The thesis introduces a novel framework that combines FL with simulated weather data collection from IoT devices, specifically WSs, whose output is FF ignition classification prediction.

b) The framework is instantiated using a FF database from the fire prone Kelowna region. The results of our framework instantiation demonstrate that the integration of IoT WSs in a FL system returns FF ignition classification prediction results comparable to a CS machine learning system.

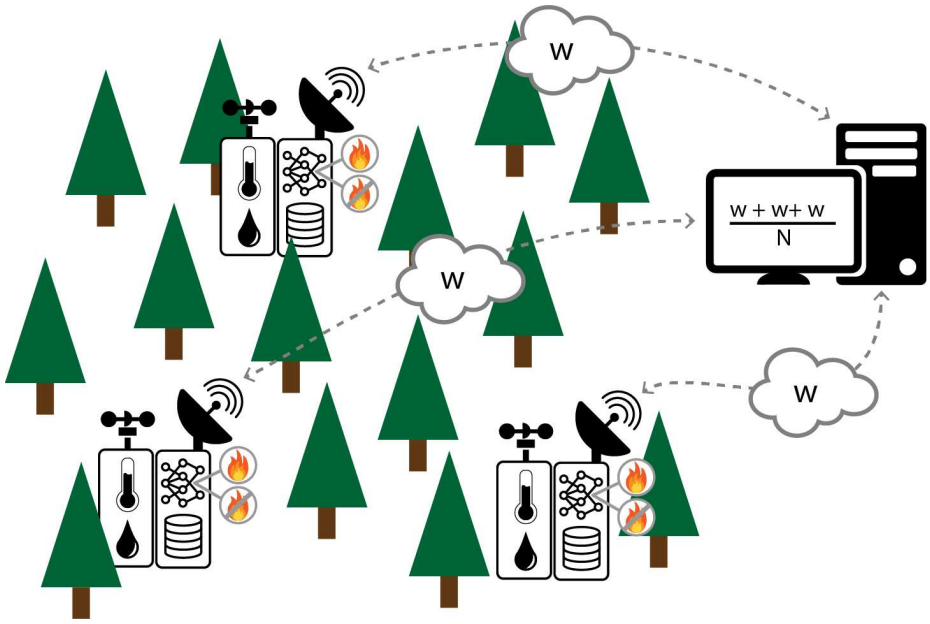


Figure 1.1: The proposed federated system with WS predicting ignition or non-ignition, and a CA that averages model weights.

c) We demonstrate that the communication energy required by a FL system for FF ignition classification prediction can be less than a CS system in some cases.

d) We demonstrate improvement in spatial precision of FF ignition classification prediction compared to a CS system.

Our contributions show that the utilization of a FL system within platforms like the CWFIS improves the system's capabilities, enabling the integration of real-time data from IoT WSs and providing localized FF ignition classification predictions comparable to a CS system. Using an imbalanced dataset with 5,008,365 non-ignitions and 45,411 ignitions, our FL system yields promising results, achieving an accuracy of approximately 0.76 and ROC-AUC of approximately 0.80.

1.3 Chapter Summary

In summary, this thesis introduces a data framework for developing forest fire ignition datasets, and a framework for using FL with IoT technologies for classification prediction of FF ignition or non-ignitions. Fig. 1.2 illustrates the overall thesis layout.

Chapter 2 provides an overview of relevant research from various fields. We examine academic papers that outline the goals and specifications of the CWFIS. We

then explore the use of IoT in smart forests and weather data collection. Next we look at papers proposing ML models for FF prediction, and finally we discuss aspects of federated learning and their relevance to this domain.

Chapter 3 presents a framework for creating datasets suitable for FF prediction research. We describe the methodology employed, and then explain the instantiation of the framework used for testing in Chapter 4.

Chapter 4 employs a dataset generated using the framework from Chapter 3 to evaluate the fitness of the models we presented in Chapter 2. First we provide details on the specific models used, then outline the methodology for assessing their fitness, and finally present the results of the evaluation.

Chapter 5 introduces a framework for applying FL to data collected from IoT weather stations (WS). We describe the instantiation of the framework and the simulation conducted to assess the fitness of the federated learning system. Next we provide an in-depth explanation of the methodology, and finally we compare the results of the simulation, specifically contrasting the outcomes of a central server (CS) system with those of a FL system.

Chapter 6 concludes the thesis with a summary of the preceding chapters and an examination of the challenges facing this work, as well as the future prospects of utilizing federated learning and IoT for FF prediction.

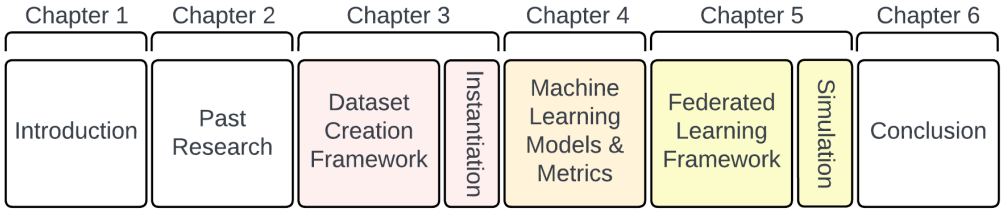


Figure 1.2: Document outline.

Chapter 2

Background and Literature Survey

This chapter introduces background information from different domains that have influenced this thesis. We first look at forest fires, then two papers produced as part of the CWFIS, then IoT and the idea of the smart forest are touched on. Next, our literature survey looks at papers from the last decade that have applied ML to the FF prediction problem. Finally the paper that originally proposed federated learning, as well as some recent FL research are looked at.

2.1 Forest Fires

As of 2021 approximately 362 million hectares (ha) of forest covered Canada, making up 9% of the world's total forested area. That same year Canada saw forest fires consume 4,307,520, 1.19% of that [6]. Forest fires, the uncontrolled burning of forest vegetation, are a positive part of healthy forest ecosystems when they consume dead plant matter and return it to the soil. However, when they threaten human communities, either through direct flame or diminished air quality, forest fire management agencies need to step in. The demand for intervention has seen a rapid increase in 2023 with the effects of climate change and extreme heat. When conditions are right, such as low moisture levels in the ground level debris of a forest, ignitions caused by lightning or humans are likely to grow beyond small sparks. The majority of fires are started by humans (52.4%), but fires started by lightning (44.4%) make up the majority of area burned (87.6%) per year [6].

2.2 Canadian Wildland Fire Information System

The CWFIS serves as a crucial platform that delivers data, analysis, and real-time information to fire management agencies and the public. It operates under the oversight of the Government of Canada through Natural Resources Canada (NRCan), which

is responsible for the management and conservation of the nation's forests. NRCan's overarching goals include sustainable development and the well-being of Canadians, encompassing the management and response to wildfires. Within this context, tools for fire monitoring and prediction, such as the CWFIS, play a vital role.

The CWFIS facilitates the sharing of information between agencies and provides valuable insights to users for fire risk assessment, aiding in prevention, resource allocation, and understanding fire behavior. Its principal focus is on FF-related data and interpretations, including fire danger ratings. These data and interpretations are sourced from various channels, including satellites and weather stations.

Since its inception, the CWFIS has undergone significant advancements. Fire management information systems have been employed in Canada since 1975, and in 1994, a national-level system displaying fire weather maps for the entire country, known as the CanFire System (now CWFIS), was established. The system consists of various modules that automatically access national data and assess the fire environment. While this paper focuses specifically on fire weather data acquisition and fire weather modeling and forecasting, other modules, such as fire behavior modeling, remain areas for future investigation.

In its original form in 1995, the CWFIS collected hourly weather data from approximately 250 weather stations across Canada, with precipitation data measured on a daily basis. The station count has since grown to over 1200. The system employed the Canadian Forest Fire Weather Index (CFFWI), an empirical model developed in 1987. This model generated calculations for each weather data collection point, providing insights into various factors related to fire, such as ignition ease, potential fuel consumption, effects of drought on forest fuels, fire spread rate, and overall fuel availability. The Fire Weather Index (FWI) served as a general indicator of fire intensity and danger. The system's outputs were utilized to create national maps, using a grid cell size of 1000 meters for the most fire-prone regions, resulting in 35 million cells each day.

There are six outputs of the FWI System. Three moisture codes and three numeric ratings of relative potential for wildland fire as seen in Fig. 2.1. The FWI is a combination of the moisture codes and numeric ratings, and is a general index of fire danger. The CFFWI does not predict fire occurrence, the estimation of the

likelihood and timing of forest fire igitions within a specific geographic area. Lawson and Armitage [25] however say that conceptually occurrence prediction can happen by applying the FWI value at a weather collection point source. For this reason we look at the FWI as context for our ignition prediction model.

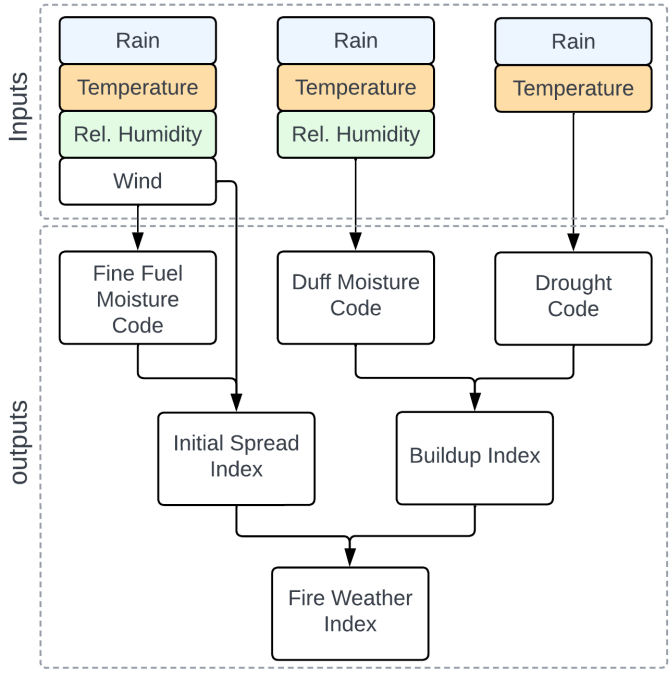


Figure 2.1: The structure of the Canadian Forest Fire Weather Index System. (adapted from [25])

As input to the FWI system four weather elements, rain, temperature, relative humidity, and wind speed, are collected manually or using automatic fire weather stations. Data is transmitted by telephone lines, digital broadband, or radio and satellite telemetry to a central network processing system. Hourly weather data is collected from weather stations and the empirical model calculations are run once a day to produce the six outputs of the FWI system. It is unclear from the literature why the calculation output is daily instead of hourly. The cell resolution of the system is 2 km with areas between weather stations using interpolated values [25].

This system is robust and has provided valuable outputs for decades. Lawson and Armitage [25] however say that a much larger network of fire weather stations is needed for accurate forecasting. Overview forecasts are unable to capture the fire weather intricacies of individual points with rapidly changing terrains or unusual

geographic and urban features. Reporting daily instead of a finer temporal resolution means the system cannot respond to points where weather changes rapidly, specifically wind.

The 1995 paper introducing the system expressed the belief that the CWFIS would assist fire management agencies in developing cost-effective and environmentally friendly strategies for forest protection. Since then, the CWFIS has evolved into a globally respected platform, providing valuable data, analysis, and real-time information to support fire management agencies and the public in their efforts to understand and respond to forest fires. This platform has served as a prototype for others across the globe and its methods appear in a diverse range of research.

The most recent version of the CWFIS includes the publication of the WGCFF-DRS. This guide provides comprehensive instructions on the proper placement and instrumentation of fire weather stations, as well as guidelines for conducting weather observations. The introduction of the guide emphasizes the crucial role that weather plays in the calculations of the Fire Weather Index (FWI) and the Canadian Forest Fire Behavior Prediction (FBP) system, both of which are essential components of the CWFIS. It also highlights the significance of weather factors that are not directly observable, such as the moisture content of debris on the forest floor. The guide underscores the importance of fuel conditions in the initiation and spread of wildfires.

While the Meteorological Service of Canada (MSC) provides mature weather services, the guide acknowledges the challenges in obtaining accurate and timely forecasts for specific weather elements, particularly in rugged terrains. To address this, the guide emphasizes the need for a specialized fire weather forecasting service and a more extensive network of fire weather stations. This expanded network would enable more detailed and precise spot forecasts tailored to specific locations, thus enhancing the efficiency and effectiveness of fire management efforts. The guide suggests that slope, elevation, and other measures of terrain could play a role in accurate forecasting.

The current version of the CWFIS recognizes the critical role of weather in fire danger rating and fire behavior prediction. The Weather Guide underscores the need for accurate and timely weather forecasts, particularly for challenging terrains, and

emphasizes the importance of a comprehensive network of fire weather stations to support efficient and effective fire management practices.

2.3 IoT and Smart Forests

Considering the common definition of IoT as a network of physical devices with sensors, software, and connectivity, the CWFIS has been utilizing IoT since its inception in the form of automatic fire weather stations. A paper presented at the 1995 Symposium on Geographic Information Systems outlined the CanFire System [26], which described a network of 250 weather stations continuously collecting data and transmitting it via satellite to a SPARC 10 Workstation located at the Northern Forestry Centre in Edmonton, Alberta.

Recognizing the significance of accurate ground-based weather data collection, the 2008 WGCFFDRS provides detailed specifications for locating and instrumenting fire weather stations. It highlights numerous benefits that automatic weather stations bring to fire weather prediction. These solar-powered stations gather data such as wind speed and direction, solar radiation, temperature, humidity, rainfall, and fuel moisture measurements. The individual weather stations and a central server coordinate the processing of data within a larger network.

The location and permanence of weather stations are of particular interest in this discussion. While there are no strict rules for placement, weather values collected within a 40 km range of a potential fire are highly reliable, whereas values beyond 160 km are considered completely unreliable. Additionally, non-permanent weather stations play a role in the CWFIS. Quick-deploy automatic stations prove useful in providing data from areas where prescribed burns are planned to prevent potential fires. These stations supply data to central models before and for weeks after a burn.

In addition to weather stations providing data to the CWFIS, forests are increasingly equipped with IoT devices [15]. These devices serve various purposes, such as monitoring forest health and disease, tracking logging activities, and measuring changes in forest structure. With the integration of inexpensive IoT devices that collect data from the forest, there is great potential to enhance the power of FF prediction models by leveraging this wealth of information.

2.4 Data

2.4.1 Data Collection

Data collection can be divided into that using satellite imaging as source, that using ground, water, and air based sensing, and research using pre-existing datasets.

Remote sensing primarily consists of satellite images, capturing a broad range of the visible and non-visible spectrum, in particular mid-infared radiation that is useful for detecting thermal anomalies. Burn scars are readily apparent within the visible spectrum of these images and can be used to determine overall burned area. Remote sensing can have line of sight problems as a result of cloud and smoke cover.

The satellite imaging most often used was from the moderate resolution imaging spectroradiometer (MODIS) instruments located on two orbiting spacecraft and made accessible through NASA curated websites like the Fire Information for Resource Management System (FIRMS). The entire earth surface is captured over one to two days, in various spectral bands and spatial resolutions [35]. The spatial resolution is finer than the data provided in our dataset, but the temporal resolution is larger and can be affected by smoke, cloud cover, and time of day. Huot [18], while noting some of these limitations like the conservative estimates of MODIS data due to darkness and cloud cover, creates a useful dataset with multiple features that goes beyond those available to ground based sensors.

Locally sensed data is collected using temperature, humidity, wind, and other sensors, often combined into a single weather station at ground level. Balloons, buoys, and ships collect data specific to their environment such as nitrogen content or water salinity. Included in local sensing are cameras mounted in static positions or attached to autonomous vehicles such as Unmanned Aerial Vehicles (UAV's). These cameras are capable of gathering the same specta as those found on satellites and are also limited by line of sight.

Research that develops a ML model based on sensor data is typically engineering based, either exploring optimal network configurations like Demin [12] and Singh [43], or the creation of new hardware like Abid [2] and Nosouhi [38]. These test their solutions using ML models for FF prediction but have moderate amounts of data for training, sometimes self-gathered, that is not representative of the scale necessary for

accurate ML modelled FF prediction.

Both local and remote sensing have limitations in geographic and time granularity, remote sensing due to the satellite's orbit, and local sensing due to physical limitations of the environment. Local sensing is particularly susceptible to destructive forces such as fire that may end its data gathering ability. In both satellite and sensor based research historic fire data most often comes from a combination of satellite data provided by systems like FIRMS and historical ground based observations. The proportion is never explicitly listed in the papers we looked at. To this end it must be acknowledged that ML models that claim to only use ground based sensors are using some degree of data from satellites and vice versa.

Research that relies on pre-existing datasets is the most numerous, and predominantly tests the viability of using existing ML models for FF prediction. A few investigate deep learning (DL) and time-series models. Authors use the Montesinho Park dataset from Portugal [14] [47] [37], data from the Punjab Forest Department [20], the SaskFire dataset [24], and the Canada National Fire Database [27] [31]. Some research also incorporates datasets from the University of California Irvine (UCI) machine learning repository [48] and Kaggle [41].

The collection details of data within pre-existing datasets is seldom discussed in the papers that use it, and surprisingly some research does not list collection methods at all.

2.4.2 Data Preprocessing

A FF ignition dataset is comprised of ignition data indicating time and geographic location of the ignition and features. Features in past papers range from the multispectrum images of remote sensing [18] to locally sensed meteorological data [37]. Both of these require preprocessing to create useable data. Remote sensing conveniently provides spectrum data on different layers. Since the images captured are limited in area, multiple images may need to be stitched together prior to sampling. To engineer a feature such as burn area the image processing necessary for edge detection must be done. Unlike sensor data, prepossessing of satellite images is necessary to turn it into usable geometric and numeric formats. The size of these images

is large and more suited to single server processing than forms of distributed computing. Some Research makes use of satellite obtained human-centric features like proximity to roads, railways, and agriculture [30]. Static observable features like human infrastructure can be valuable for training but it is worth noting that though human activity is responsible for approximately 52% of FF ignitions, the 44% started by lightning are responsible for approximately 87% of the area burned every year [6]. Semi-static features like vegetation that impact both human and lightning caused ignitions [28] may prove to be more useful than human infrastructure features, especially since existing vegetation datasets that preprocess images into geometry are readily available. When existing datasets have been used an assumption of the fitness of these are made. Some preprocessing is done, most often normalization, but seldom reflect the uniquely unbalanced nature of FF ignition data. The use of synthetic minority over-sampling technique (SMOTE) in some research is an exception [28]. Rows where data is missing are usually removed, with no attempts to generate replacements. Resolution does not come up, with most of the papers we looked at relying on the highest resolution available. The spatial resolution in satellite images maybe quite high and consistent, unlike locally sensed data where weather stations are unevenly distributed with in-between data being generated by trusted forecasting algorithms.

2.5 Machine Learning Models

For the purpose of our research, we define FF prediction as the classification of whether an ignition will occur within a specified time increment in a defined geographic area. Previous studies that adopt this classification approach employ various ML algorithms and metrics to model and assess the relationship between features and historical ignition data. The models primarily fall into the categories of time-series models, neural networks (NN), and the most prevalent category, decision trees. This thesis primarily looks at the models found in Table 2.1.

To examine the presence of FF prediction in ML research we did a search on IEEE Xplore using the term 'Forest Fire Prediction' In total, 38 papers met this criteria within the period from 2014 to 2022. A trend of increasing publications can be seen in Fig. 2.2. In the following sections we look at these publications based on their

approach to data, machine learning models used, and evaluation metrics used.

Table 2.1: Models used in this thesis. (in alphabetical order)

model	description
Bag Tree	Bagging stands for bootstrap aggregation. It is an ensemble of decision trees where each tree generates a prediction using a subset of the total data and all the predictions are averaged. [22]
Decision Tree	A decsion tree model simply applies an if/then approach to the values in a dataset, creating a binary tree of set depth. Many of the other models in this chart are built off of different approaches to multiple decsion trees.
Multilayer Perceptron (MLP)	A feed forward neural network, the multilayer perceptron is made up of an input layer, an output layer, and a least one hidden layer. To learn data, flows in only one direction. Back propagation is used to adjust weights between perceptrons based on the error between expected output and actual output.
K Nearest Neighbor (KNN)	In this model predictions are made based on how similar a new sample is to k number of points in the training data set.
Logistic Regression	This model is considered the 'hello world' of machine learning. An equation is used to calculate the log odds probability that an observation containing specific variable values will be 1, based on a probability threshold. [44]
Random Forest (RF)	A Random Forest model combines multiple decision trees that each use a sample data set made up of randomly chosen rows and features from the original dataset. In the case of classification the majority vote from all the trees is the prediction. [22]
Support Vector Machine (SVM)	Support Vector Machine tries to find a hyperplane or line using a statistical framework that best separates training data into two classes. New data is assigned to one of the two classes depending on where it falls.
Long Short-Term Memory (LSTM)	Long Short-Term Memory models are a type of recurrent neural network (RNN). They selectively retain or forget data over sequences, enabling them to capture temporal patterns and dependencies.
XGBoost tree	A boosting model starts with a first model and then each subsequent model is built to correct the errors of the previous model. This process continues until the training data set is predicted correctly or the maximum number of models is reached. [22] An extreme gradient boosted model assigns weights to variables which are modified in each subsequent model.

Random Forest (RF) and Support Vector Machine (SVM) ML models have been frequently employed in the literature we reviewed, appearing in 9 and 8 studies, respectively. These models are well-suited for the classification task presented by FF prediction and often serve as the initial choices tested before researchers explore more complex decision tree-based models. For instance, some research progresses from RF and SVM to gradient boosting classification [17], while some transitions from RF to various complex boosted trees before attempting a shallow NN [16].

In terms of NN models, we observed a lack of standardization in the literature we surveyed. The recent increase in processing power provided by GPUs may partially account for this. At a superficial level, NN models appear to be more suitable for the FF prediction task, particularly in scenarios related to image analysis, such as convolutional NNs, or when large training sets are necessary to capture the infrequency of FF events. Decision trees tend to experience efficiency challenges with increasing input sizes. Recurrent NN models, such as LSTM, have shown promise in some research [36] [31] [28] [27].

Overall, the existing literature reflects a diverse range of ML models and techniques employed for FF prediction. While RF and SVM models are commonly used as initial choices, the exploration of more sophisticated decision tree models and NN architectures is becoming increasingly prevalent. The potential of recurrent NNs, particularly LSTM, has shown promise in capturing temporal dependencies and improving prediction accuracy. However, further research is necessary to establish standardization and identify the most effective models and approaches for FF prediction, considering factors such as data characteristics, computational resources, and the specific requirements of the prediction task.

2.5.1 Evaluation Metrics and Results

The goal of evaluation metrics are to measure how well predicted outcomes match observed outcomes. The choice of metric is critical when considering highly unbalanced data like that for FF ignitions. If a region is divided into 2D grid cells and the time for each cell is divided into increments such as hours the fitness of a ML model's ability to classify whether an ignition will happen can be measured based on counts of the following:

- **True Negative (TN) or True non-ignition** — The correct classification of a cell/time combination where no ignition takes place. A high number is desirable.

- **False Negative (FN) or False non-ignition** — The incorrect classification of a cell/time combination where an ignition does take place. The cell at the time is thought to have no ignition when in fact it does. This is the most

dangerous situation and the smallest number possible is desirable.

- **True Positive (TP) or True ignition** — The correct classification of a cell/time combination where an ignition does take place. The ideal situation is to identify every possible ignition.
- **False Positive (FP) or False ignition** — The incorrect classification of a cell/time combination where no ignition takes place. The cell at the time is thought to have an ignition when in fact there is none. This situation results in misallocation of fire fighting resources so a lower number is desirable.

Table 2.2: Commonly used metrics to assess machine learning models.

metric	description	other
accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Ratio of predictions the model got right. Doesn't work well if data is unbalanced.
precision	$\frac{TP}{TP+FP}$	Ratio of positive predictions the model got right. Doesn't work well if data is unbalanced.
recall	$\frac{TP}{TP+FN}$	Ratio of true positive predictions the model got right. The inclusion of FN makes this good for forest fire prediction and should be as high as possible.
F1-score	$2 \times \frac{precision \times recall}{precision+recall}$	Better for classifying true negatives in unbalanced data than accuracy is.
sensitivity	$\frac{TP}{TP+FN}$	This is another name for recall. It is plotted agains't specificity in ROC.
specificity	$\frac{FP}{FP+TN}$	This is the rate of false positives. Ideally this is low in forest fire prediction.
ROC	$\frac{sensitivity}{specificity}$	Gives the resulting curve when sensitivity and specificity are plotted agains't each other. To get a single value a specific sensitivity or specificity must be chosen.
ROC-AUC	$\frac{sensitivity}{specificity}$	Gives the area under the resulting curve when sensitivity and specificity are plotted agains't each other. It gives a single value.

Common metrics that use TP, FP, TN, and FN can be seen in Table 2.2. A common measure, with the highest count in the FF research we looked at, was accuracy. Also

high in the research was precision. A small proportion of the research we looked at used Sensitivity and Specificity [37] [20] [23]. Approximately 50% of the research we looked at used Recall and the area under the receiver operating characteristic curve (ROC-AUC) which is a measure of the ratio *Sensitivity/(1 - Specificity)*. Only four of the papers we looked at used F1-score [28] [23] [19] [4].

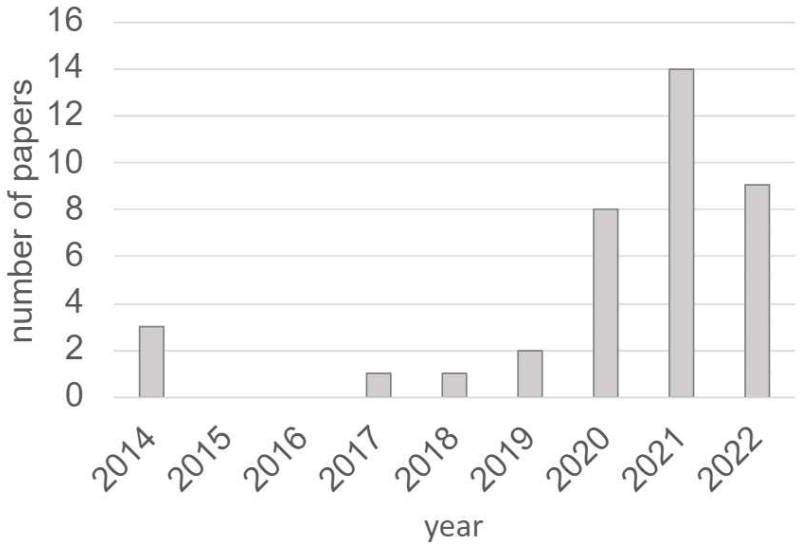


Figure 2.2: The number of papers within our search parameter in IEEE Xplore by year.

2.6 Federated Learning

Introduced in 2017 [29], federated learning involves training individual clients on their respective datasets, with only model updates shared with the global model. The privacy advantages of this approach have been extensively studied. While privacy may not initially appear crucial for weather data, [33] in an upcoming publication propose a multi-modal federated learning system for FF prediction, where models trained on isolated data sources are utilized. In a previous work, [32] employed a federated learning approach to predict FF probability and severity; however, unlike our study, their work did not incorporate spatial or temporal input. Apart from the aforementioned papers, FF prediction has limited representation in federated learning research. Federated learning is more commonly employed in applications where a generalized model is required, but data must remain local and secure, such

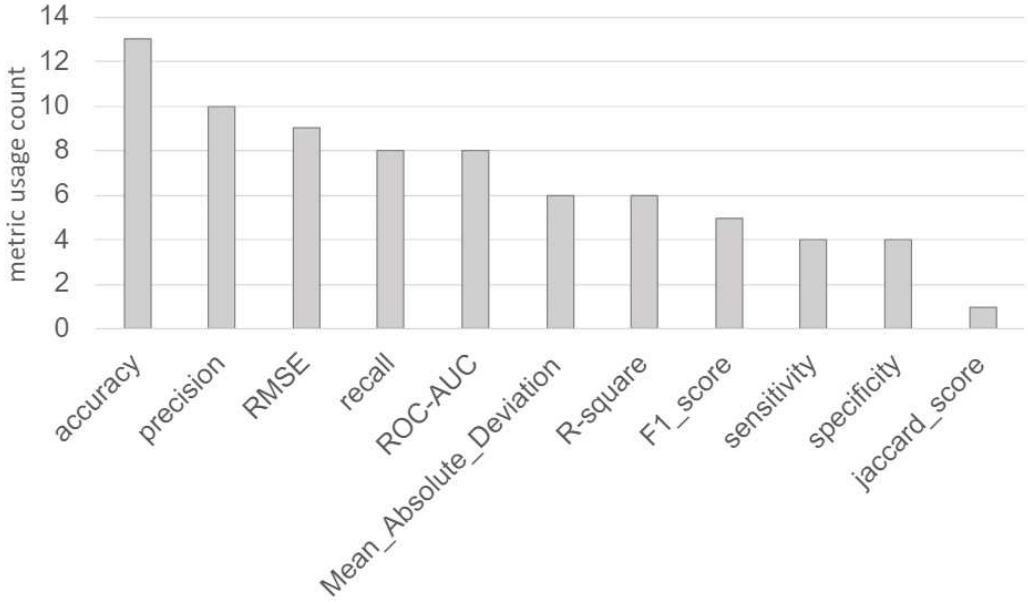


Figure 2.3: The occurence of metrics in papers reviewed.

as in mobile devices and autonomous vehicles.

2.7 Research Gap and Motivation

Given the escalating severity of forest fires, the lengthening fire weather seasons, and the heightened threat to communities, it is imperative for researchers to explore all possible approaches to mitigate this hazard. The limited presence of federated learning in FF prediction research is difficult to justify, considering its capability to facilitate efficient data sharing and its potential to enhance spatial and temporal predictions. Closing this research gap and harnessing the benefits of federated learning can significantly contribute to the advancement of FF prediction methods and ultimately aid in effective fire management and prevention.

2.8 Chapter Summary

In this chapter, we have examined previous research from various domains that have influenced the content of this thesis. Specifically, we have explored the CWFIS, IoT technologies, and the concept of smart forests. Additionally, we have reviewed relevant studies from the past decade that have applied ML techniques to address the FF prediction problem. Moreover, we have studied the original paper that introduced

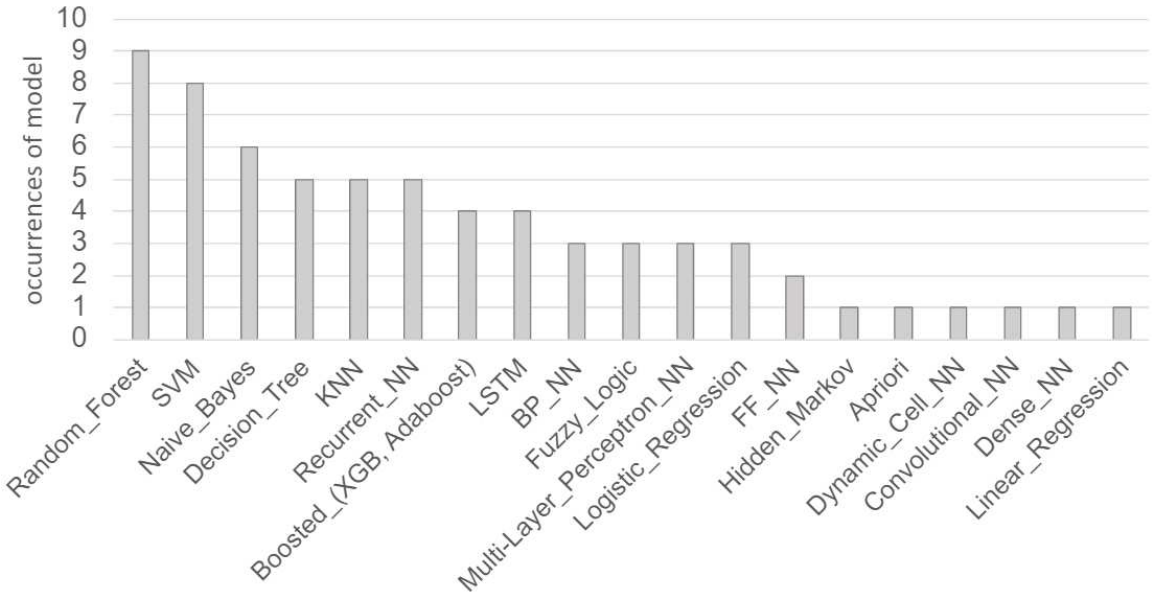


Figure 2.4: The occurence of model types in papers reviewed.

the concept of FL, as well as recent advancements in FL research.

Moving forward, the subsequent chapter will delve into a comprehensive framework designed to create datasets suitable for training ML models in FF ignition prediction.

Chapter 3

Dataset Creation Framework

In this section, we present the dataset creation framework and provide a detailed description of its implementation. Specifically, we focus on instantiating the framework to generate a dataset aimed at predicting forest fire FF ignitions in BC, Canada. To assess the effectiveness of the framework, we develop a novel FF prediction dataset and elaborate on the data sources utilized for its instantiation.

Algorithm 1 Aggregate

```
Create dataframe predictor_features
for years = start_year, start_year + 1, ..., end_year
    for day = 1, 2, ..., total_days
        for location = 1, 2, ..., total_locations
            for feature = 1, 2, ..., total_features
                Get each geotif representing the feature and day
                Turn each geotif into a vector
                for vector_element = 1, 2, ..., total_elements
                    Create temp_row for vector_element
                end for
            end for
            Merge temp_row for current location-date with predictor_features
        end for
    end for
end for
Add ignition column to predictor_features with default zero
Add fireID column to predictor_features with default zero
Add fireCause column to predictor_features with default empty string
ignitions = historic_ignitions.shp
for ignition = 1, 2, ..., total_ignitions
    if ignition.location = predictor_features.location
        if ignition.date = predictor_features.date
            Set ignition to 1
            Set fireID to ignitions.fireID
            Set fireCause to ignitions.fireCause
        end if
    end if
end for
Write out predictor_features_with_ignitions.csv
```

3.1 Framework

The framework begins with input from three sources as indicated by the top row of Fig. 5.1. Weather (weather source) and ignition sources (ignition source) are expected to have coordinate and date data, ignition data may also include how the ignition was initiated (i.e. lightning), the resulting burn area, and the end date of the resulting fire. Weather data typically includes air temperature, relative humidity, wind direction, surface pressure, and precipitation. The temporal and spatial resolution of these sources determine the resolution of the final dataset. Pseudocode for the transformation of features into rows are shown in Algorithm 1, lines 1-15. Pseudocode for the addition of ignitions to the rows is shown in Algorithm 1, lines 16-26. The third input is a boundary (boundary source) of the area of interest.

Historical weather data is available from regional government agencies such as Environment Canada and the American National and Oceanic Atmospheric Administration. Formats available include csv, GeoJSON, and GeoTIFF. The goal of our framework is to use historic weather data that is available for all regions globally. Climate reanalysis data, where data is generated from existing climate models, provides a reliable source that is well validated [10], available for many years, and provides samples from all regions equally across the globe. Existing reanalysis models may be directly incorporated into the framework. Due to its ease of use and previously mentioned validation we, however, chose to use the web dataset generation tool provided in the Copernicus Project by the European Center for Medium Weather Forecasting (ECMWF) [34].

Historical FF ignition data (ign_src) is not consistently available. Some regions, like BC, Canada, offer public datasets with igniton dates and times. Some regions like Manitoba (MB), Canada do not. Some datasets have ignition dates recorded in multiple formats. For instance the province of Alberta (AB), Canada, offers 4 datasets each with a different time format and recording period [39]. This inconsistency can be attributed to the recording of ignitions based on human memory and observation prior to the use of satellite ignition spotting made available through systems like Nasa's FIRMS [35]. The largest database containing dates and times for ignitions is the "Combined wildfire datasets for the United States and certain territories, 1878-2019" available from U.S. Geological Survey [45]. To use past historical data from

regions that include ignition dates but not times we randomly select times based on the probability of ignition calculated from datasets that do have dates and times.

A boundary (bnd_src) is necessary within the framework to provide limits for both the weather data and the FF ignition data. Weather data can be available only in rectangular image format which reflects its satellite imaging origin. Historical FF ignition data is any FF data that has the start time (ignition) of a FF at a particular location at anytime prior to the current moment. These may be human observed and recorded or satellite imagery that contains ground level heat data. Human observation and entry errors which place ignitions outside the requested region's boundaries are present in some of this data. For these reasons a boundary (boundary source) is necessary to mask both weather sources and ignition sources as seen in box 1 of Fig. 5.1. The process of masking involves generating a rectangular black and white image where white represents the region of interest. The pixel value of weather data represented as an image, such as temperature, is multiplied by the masking image when converting image data to numeric data. Any data falling outside the region of interest becomes 0 and data within retains its original value.

Once the input data has been masked, the weather data and the historical FF ignition data are combined, as in box 2 of Fig. 5.1. A sample row demonstrating this combination can be seen in Appendix B. During the combination pre-processing may need to be done to align the formats of the two sources. For instance, reanalysis data from the Copernicus Project is provided in Network Common Data Form (NetCDF) and needs conversion from image to numerical data. This is done by reading an image's metadata containing unit information and sampling the pixel value at a location. Date and time formats also need to be standardized. When instantiating the framework we converted all temporal data to the 64-bit Network Time Protocol (NTP) which writes date and time as seconds after 1900.

Our goal was to create datasets in the .csv format with a row for each location and time. Every row contained coordinates, feature values, and a binary column representing whether an ignition occurred for that location and time. This format is represented in our framework diagram, Fig. 5.1, as 'final data'.

3.2 Instatiation

To test our framework we instantiated it in several regions. In this thesis we describe the instantiation of the region that covers the southern third of BC, Canada. The first two rows of this instantiation can be seen in Appendix B. An initial ad hoc survey of data sets showed that while weather data and wildfire ignition data was available as far back as the late 19th and early 20th centuries for BC, reliable consistent time series data was not available until after 1950. We chose to look for data available for a period of 41 years, from 1980 to 2020. This time period contained a substantial number of wildfires and was long enough to be representative of natural cycles like prevalence of lightning strikes. Next we look at the specific data sources used when instantiating the data collection framework. Finally we present the specifics of aggregating the collected data.

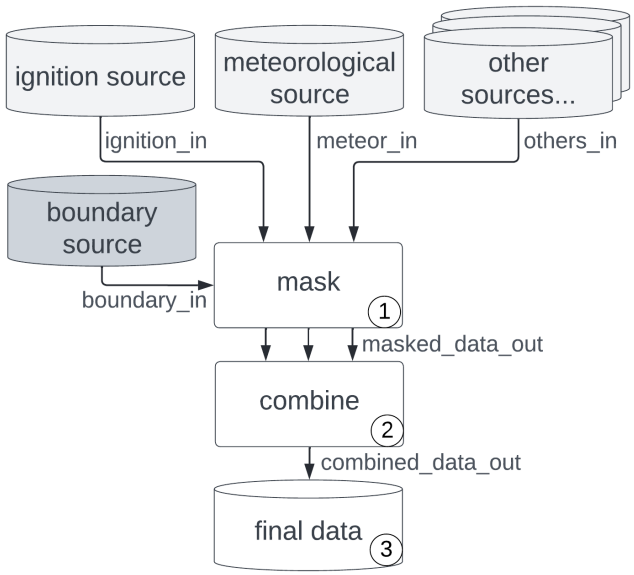


Figure 3.1: Data collection framework diagram.

3.2.1 ERA5-Land Data

The principal data used to gather predictor features for this research was ERA5-Land [34], represented in our data collection flow diagram, Fig. 3.2, as 'weather source'. ERA5-Land data is delivered as multi-layer NetCDF files containing gridded

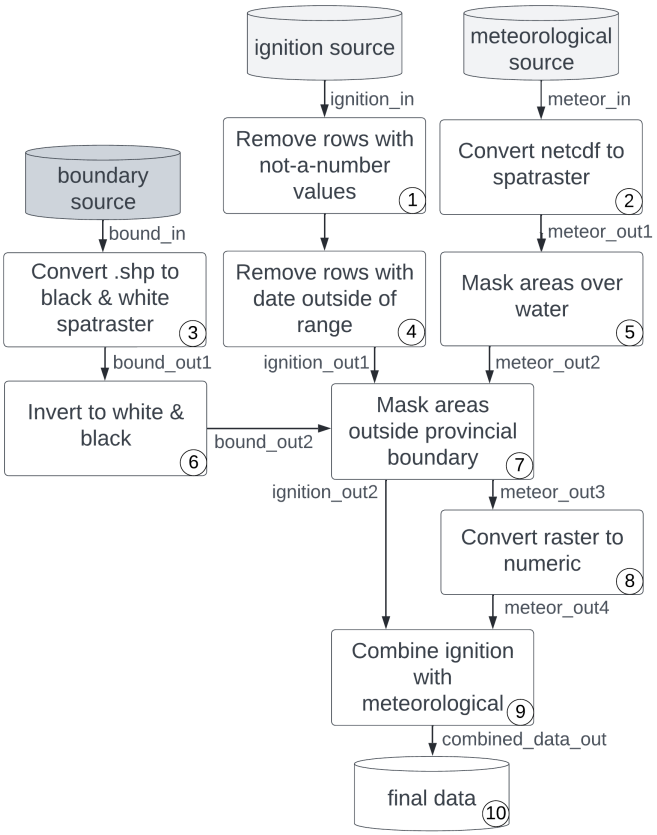


Figure 3.2: Instantiation of framework diagram.

data that is projected onto a regular longitude and latitude grid. A sample can be seen in Fig. 3.3 The horizontal resolution is 0.1° x 0.1°, or approximately 11 km^2 depending on the longitude. Vertically, measurements extend from 2 m above the earth's surface to a soil depth of 289 cm. A temporal resolution of 1 hour is available but we chose to only request measurements at 12 noon. This was to limit data points and to follow the lead of the widely used Canadian Fire Weather Index (FWI).

Another way to limit data points requested from ERA5-Land was to limit the months we looked at. An analysis of wildfire ignitions in BC revealed that less than 1% of ignitions occurred in the months from October to April. For this reason we only requested reanalysis data from May through September.

Copernicus limits the number of downloadable data points per session. To meet this limitation we downloaded and stored single years, May through September, from 1980 to 2020. The boundaries -139.1° west, -114.0° east, 60.0° north, and 48.5° south were used. This area encompassed the complete province of BC with some buffer area

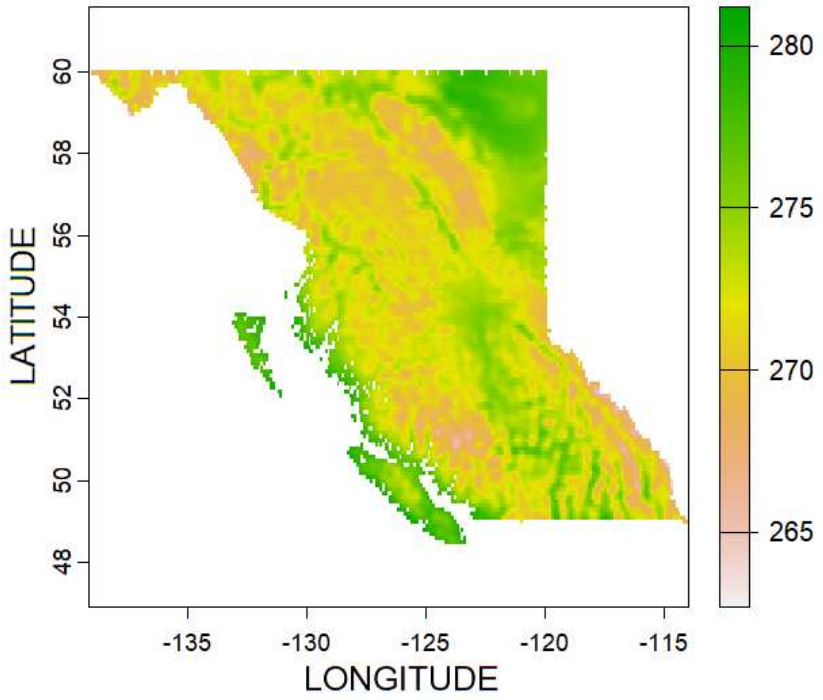


Figure 3.3: A NetCDF plot representing temperature in Kelvin for BC on May 01, 2010. Areas over substantial bodies of water have been removed.

extending beyond the provincial boundary.

Our downloaded result was 41 files in the NetCDF format. Each file contained 3496 GeoTIFFs, image layers, with geographical information. Each layer represented a feature on a single day of the year. For instance, temperature data for April 01, 1980 for our area of interest could be found on one layer. To use the layers within the R language these were converted to the R internal format of spatraster as seen in box 2 of Fig. 3.2.

3.2.2 Fire Incident Locations

The outcome features were obtained from fire incident locations provided by the government of BC [13], and represented in our data collection flow diagram, Fig. 3.2, as 'ignition source'. This dataset provided historical wildfire incident locations for BC prior to the current fire season. All ignitions from various sources that are tracked by the BC Wildfire Service are included. Data is presented in the commonly used

GIS format .shp. It is unclear whether the incident locations are from ground observations or satellite. The presence of fire incident locations listed outside of provincial boundaries indicates that there may be data entry errors or other collection problems that require correction. Fire ignition locations originating as far back as 1950

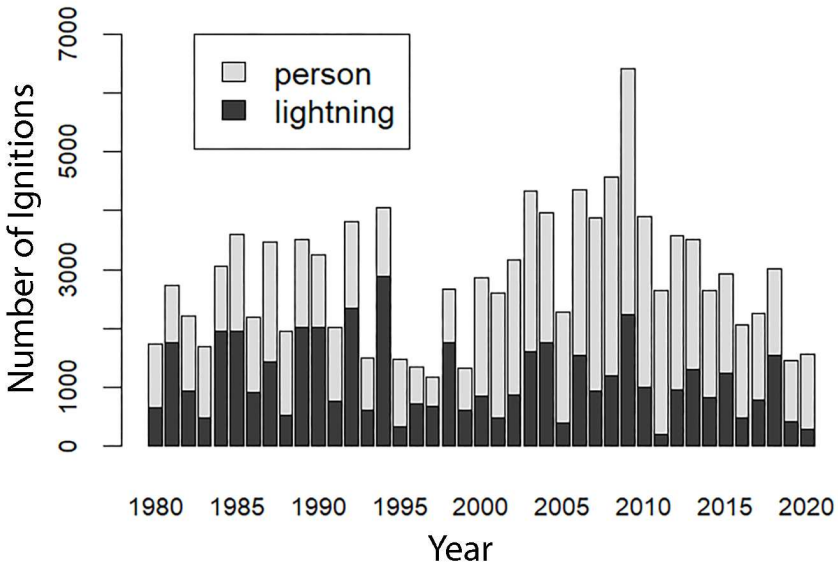


Figure 3.4: Number of ignitions by year, and cause for BC.

were available from this data. Some ignitions had locations outside of the provincial boundaries. These made up approximately 0.2% of the entire dataset. We removed these ignitions as well as any ignitions prior to 1980 and after 2020, represented in box 1 of Fig. 3.2. Rows with not-a-number (NAN) information in the longitude, latitude, ignition date, and ignition cause were also removed. This is represented in box 4 of Fig. 3.2. After cleaning the downloaded fire ignitions dataset sample from 1980 through 2020 we performed an analysis using the R language's internal summary function. This function provides means and standard deviations for data columns. Temporally the data shows a varied number of ignitions per year with the overall mean being 1132 ignitions per year, with the max in one year being 2888 and the min in one year being 213. As seen in Fig. 3.5 June and July have by far the most ignitions. There is no discernible pattern in the number of ignitions per year with the exception that there is an increase over the 41 year period looked at. If the number of ignitions per year is separated by cause two different patterns emerge as seen in

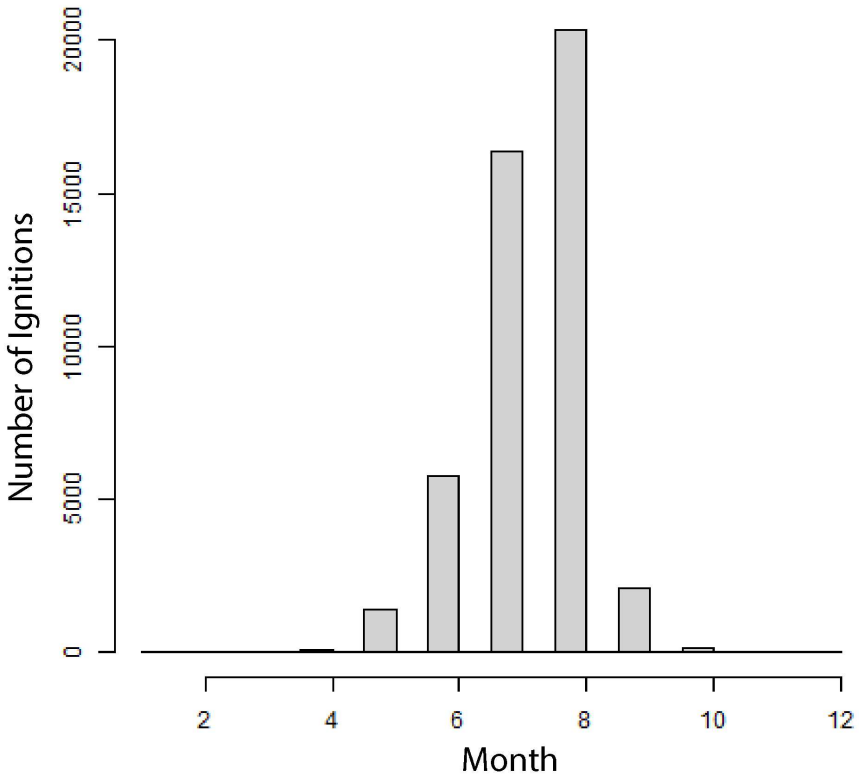


Figure 3.5: Number of ignitions by month.

Fig. 3.4. Person started ignitions appear to happen in continuous streaks, from 2000 to 2010 for instance. Lightning caused ignitions appear to happen cyclically. The number rises and falls in a repeating pattern, with a complete cycle taking 8 years. Over the 41 year period we looked at there are 8 cycles.

Spatially ignitions cluster in the lower third of British Columbia BC. This can be seen in Fig. 3.6. When separated by cause, as seen in Fig. 3.7, human started ignitions appear more dense in the western half of the lower mainland, and lightning started ignitions appear more dense in the eastern half.

3.2.3 Boundary Data

To limit the two previous sets of data to within the provincial boundaries of BC, the provincial outline provided by the Government of Canada was used [7]. This boundary is represented in our flow diagram as 'boundary source' in Fig. 3.2. Download options chosen from the 2021 Census Boundary web page were: Language-english, Type-cartographic boundaries, Administrative Boundaries-provinces/territories, and

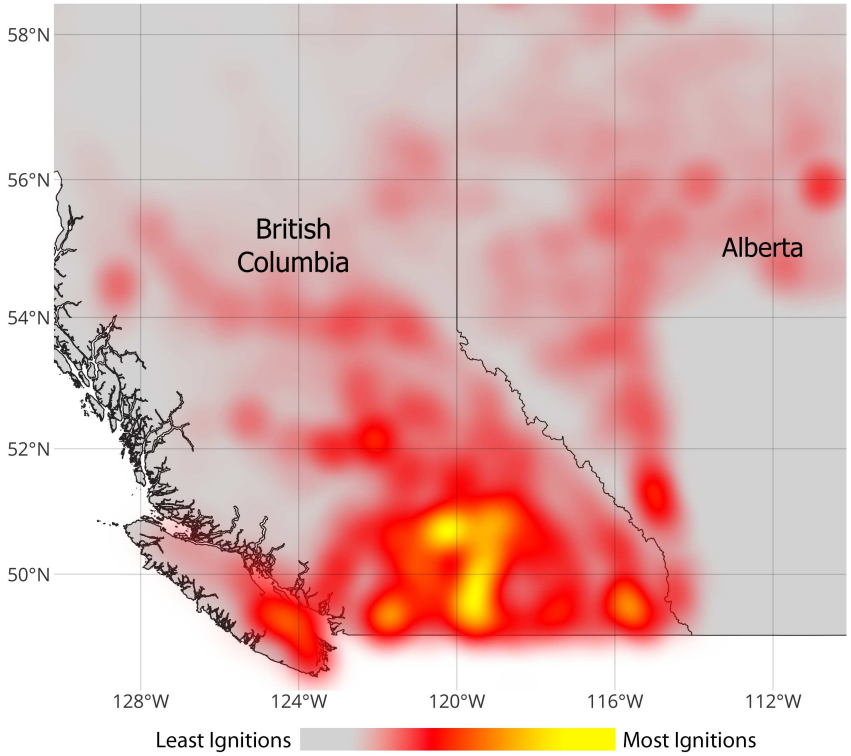


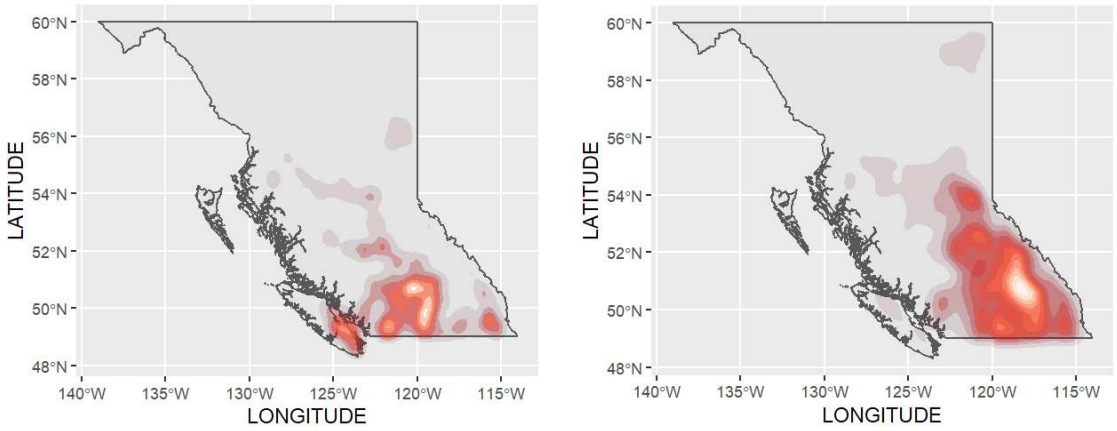
Figure 3.6: Heatmap of ignitions from 1980 to 2020.

Format-Shapefile.

The polygon for BC was extracted and converted to a black and white spatraster object using the terra package in R, represented in box 2 of Fig. 3.2. The terra package contains methods for spatial data analysis [1]. The spatraster was inverted so that areas outside the region of interest were black and areas within the region of interest were white, represented in box 3 of Fig. 3.2. The spatraster object was then used to mask weather and fire data, represented in box 7 of Fig. 3.2.

3.2.4 Aggregation

We combined the ERA5-Land reanalysis data with the BC historical fire incident locations data as seen in Algorithm 1. Feature GeoTIFFs contained in the NetCDF files from Copernicus were turned into rows, where each row represented a location and day. To these rows FF ignition data was then added. Columns, as seen in Table 3.1 created were location, year, day of the year, land based weather features, a binary indicator of ignition or no ignition, the resulting wildfire ID if applicable, and



(a) Person started ignitions.

(b) Lightning started ignitions.

Figure 3.7: Ignition density heatmaps from 1980-2020.

the ignition cause if applicable.

Table 3.1: Headings of dataset combining weather data and FF ignition data.

GridID	Lon	Lat	Features 1, 2, 3,..., n.	DOY	Year	FireID	Ign	Ttl Ign	Ttl Person	Ttl Lightning	Ttl Unknown
-	-	-	-	-	-	-	-	-	-	-	-

3.2.5 Final Datasets

Three datasets were produced over the course of this thesis. An overview can be seen in Table 3.2. Features and summaries of the datasets can be seen in Appendix A, C, D, E, F.

Table 3.2: Summary of Instantiated Datasets.

Dataset	Location	Total Rows	Features	Classes	non-ignitions	ignitions
High Prarie	AB, Canada	79,387,152	25	1	79,369,847	17,305
Kelowna	BC, Canada	68,437,200	25	1	68,385,282	51,918
Lower BC	BC, Canada	25,728,210	23	1	25,662,473	65,737

3.3 Chapter Summary

In this section, we have introduced a dataset creation framework and extensively discussed its implementation. Our primary focus was on utilizing the framework to generate a dataset specifically designed for predicting forest fire FF ignitions in BC, Canada. Through this process, we successfully developed a novel FF prediction dataset, and we provided comprehensive insights into the data sources employed during its instantiation.

In the subsequent chapter, we will delve into the realm of machine learning models and metrics, particularly in the context of FF prediction. We will explore various machine learning algorithms and evaluation metrics, considering their applicability and effectiveness in addressing the FF prediction challenge.

Chapter 4

Machine Learning Models & Metrics

This chapter focuses on evaluating the fitness of the dataset obtained through the instantiation of the data collection framework by modeling it. The term "fitness" in this context refers to the degree of suitability of the produced dataset for the specific task of FF ignition prediction. To achieve this objective, we train ML models that have been previously utilized in FF prediction research using the collected data. Subsequently, we employ established metrics from prior studies to assess the performance of the trained models, thereby evaluating their effectiveness in predicting FF ignition. Furthermore, we critically evaluate the suitability of these metrics themselves for effectively measuring the performance of the models in the context of FF prediction.

Efficiently representing the relationships between features and targets is crucial in data modeling. In our analysis, we focus on two key aspects: sampling and feature correlation. We demonstrate that by sampling a region based on ignition frequency, it is possible to significantly reduce the database size without disrupting the ratio of ignition to non-ignition points. Additionally, by understanding the correlation between features, redundant ones can be identified and eliminated. This approach aims to achieve the most efficient representation of the data, using a model that yields good results with the minimum number of features.

From a conceptual perspective, we illustrate in the following sections that our data can be effectively modeled using this efficient approach. Moreover, from a practical standpoint, the efficiencies gained through sampling and feature correlation demonstrate that our data can be scaled and adapted to systems with limited processing resources.

4.1 Modelling Workflow

The R language and RStudio provide a comprehensive development environment, encompassing various libraries suitable for both ML model development and statistical analysis. The active and supportive community surrounding R promotes result reproducibility, facilitates collaboration, and ensures high code transparency. Therefore, in this study, we chose to evaluate the instantiated dataset using R and implemented the workflow depicted in Fig. 4.1. To leverage the advantages of R, we followed the principles of the Tidyverse paradigm, which is a collection of R packages that offer a consistent framework for tasks related to data manipulation, exploration, and visualization. Specifically, our workflow was based on the "Screening Many Models" chapter of the book *Tidy Modelling with R* [21].

Initially, a random subset of years was selected from the dataset and loaded into memory. Any non-feature columns were removed (Box 1 in Fig. 4.1). Subsequently, this sample was divided into train and test sets using a 3:1 ratio (Box 2 in Fig. 4.1). The training set consisted of data from 1981 to 2010, while the test set encompassed the years 2011 to 2020. The training set underwent resampling with 5 repeats of 10-fold cross-validation. The data split for training, validation, and testing was arbitrarily set at 70%, 15%, and 15%, respectively. The folds were stratified based on the year (Box 4 in Fig. 4.1). To address the significant data imbalance between non-ignition and ignition cells, the training set was downsampled prior to creating the folds (Box 3 in Fig. 4.1). The evaluation of different downsampling ratios revealed that the highest ROC-AUC and Recall values could be achieved by using a downsampling ratio of 1:1. The results of downsampling ratios ranging from 1:1 to 5:1 were considered.

Following the Tidyverse paradigm, separate recipes were created for each desired model (Box 5 in Fig. 4.1). Additionally, preprocessing recipes were developed to accommodate model-specific requirements, such as normalization (Box 6 in Fig. 4.1). Each preprocessing recipe incorporated a function to remove highly correlated features above a predetermined threshold of 0.7. Rows containing non-numeric values were also eliminated. The model and preprocessing recipes were combined into Tidyverse workflows (Box 7 in Fig. 4.1).

The models underwent tuning using a 10 x 10 grid, applying the same function

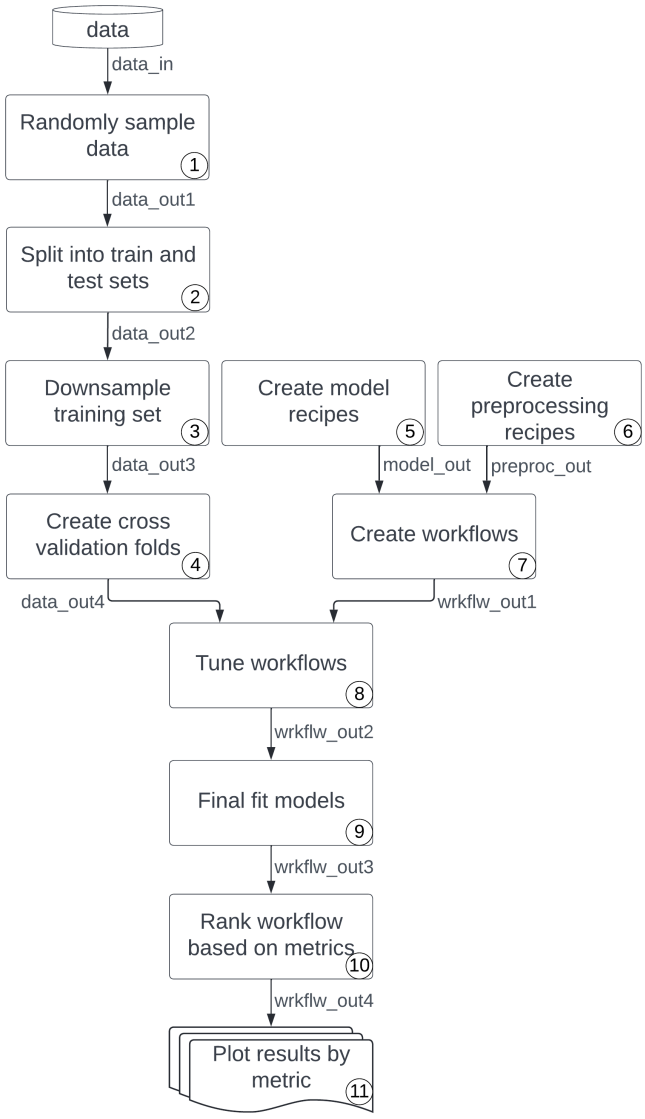


Figure 4.1: Diagrammatic representation of the initial modelling flow.

to each workflow (Box 8 in Fig. 4.1). Once the optimal tuning parameters were determined or the grid limit was reached, a final fit was performed on each workflow (Box 9 in Fig.4.1). The models were then ranked based on specified metrics, and the results of each ranking were plotted (Box 10 and 11 in Fig. 4.1).

4.2 Model Results

Examining the model results for ROC-AUC in Fig. 4.2, it is evident that most models cluster between the range of 0.76 and 0.82, with RF ranking first. However, Logistic Regression ranks as the poorest performing model with a value of approximately 0.5. This outcome could be attributed to the imbalance present in the test set and the inadequate tuning of the prediction probability threshold for assigning outcomes to classes.

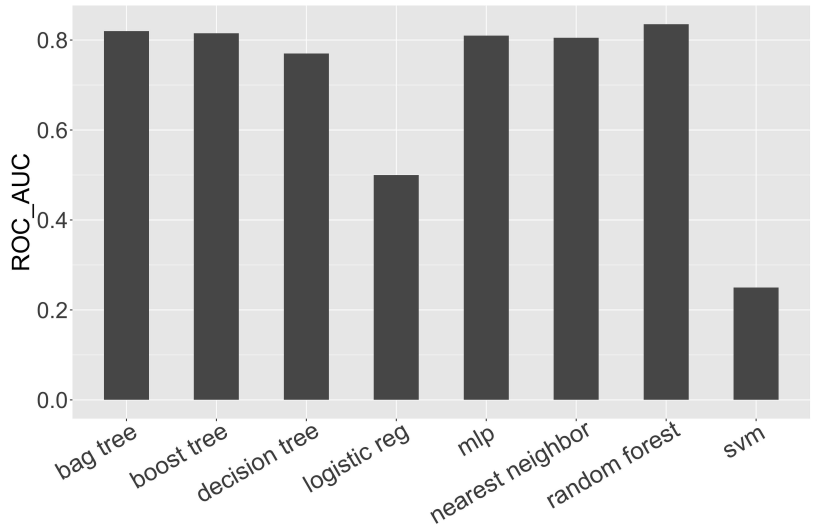


Figure 4.2: ROC-AUC score of initial 8 models.

Considering Recall, RF also emerges as the top-performing model, as seen in Fig. 4.3. Recall measures the number of correctly predicted true ignitions divided by the sum of the total true ignitions and false non-ignitions. As false non-ignitions pose the highest risk by potentially leading to unanticipated FF, this metric holds significant importance in our research. The clustering of most models between approximately 0.67 and 0.72 suggests that additional preprocessing of the training data or further model tuning is necessary. Both SVM and Logistic Regression exhibit poor performance in terms of Recall. Potential reasons for SVM's underperformance include inadequate hyperparameter tuning, inappropriate selection of the SVM kernel, the large size of the dataset, or the impact of normalizing the training data. Investigating these factors is left for future work. Notably, both Logistic Regression and SVM exhibit extremely large standard errors in the Recall metric, indicating that the

models do not converge into reliable models.

The Sensitivity metric, which represents the proportion of correctly predicted outcomes, exhibits similar results to both ROC-AUC and Recall, with RF ranking first and SVM and Logistic Regression ranking last (with large standard errors), as depicted in Fig.4.3.

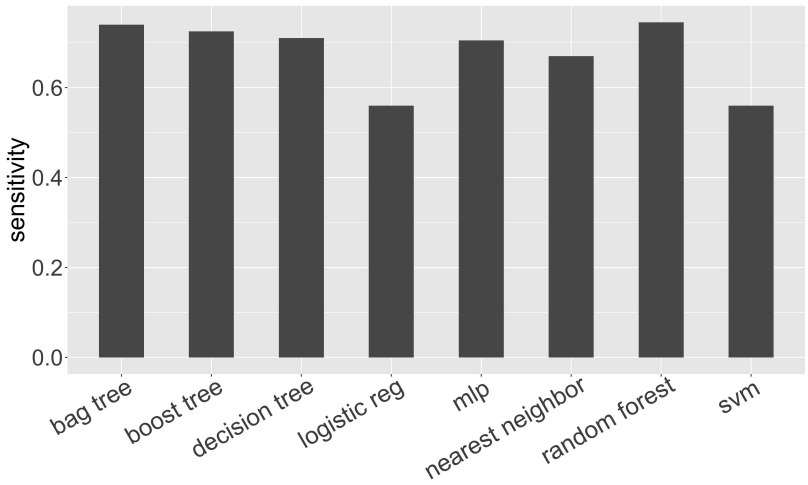


Figure 4.3: Sensitivity (also Recall) score of initial 8 models.

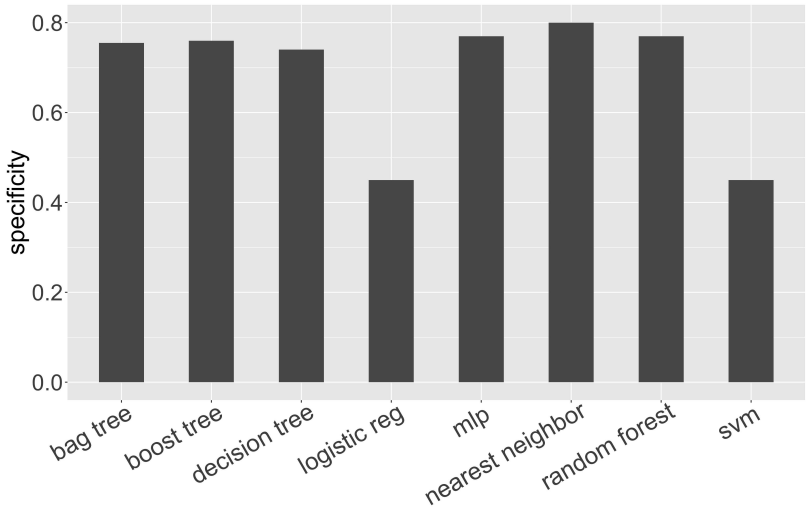


Figure 4.4: Specificity score of initial 8 models.

The results for Specificity, as shown in Fig. 4.4, present a different perspective compared to other metrics. KNN ranks the highest, closely followed by MLP and RF. While reducing false non-ignitions is more crucial than reducing false ignition

predictions in terms of resource allocation, the choice of preferred models may vary depending on the intended usage.

4.3 Sampling

Analysis of heat maps depicting historic BC FF ignition locations revealed that a sample from the lower third of the province encompasses the majority of ignitions. Gradually decreasing the latitude showed minimal changes in the ratio of ignitions to non-ignitions, which can be attributed to the large number of samples utilized (Fig. 4.5). The overall reduction in the ratio between latitudes of 50° and 60° was approximately 0.005. Considering the negligible impact of latitude on the ignition to non-ignition ratio, a sample for modeling was selected between longitude −115° and −124° and latitude 49° and 53° to reduce the volume of data. This reduction was combined with the previously mentioned reductions in the years and months sampled (Section 3.2.1), resulting in a reduction of 75,108,094 fire cells (ignition + non-ignition) in the dataset (Table 4.1).

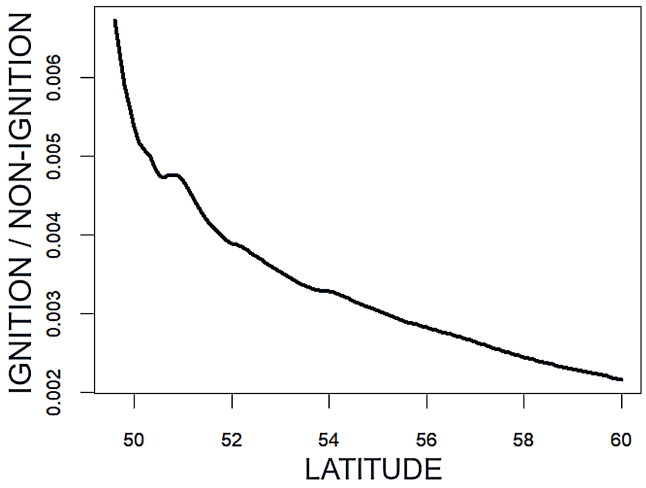


Figure 4.5: Ratio of non-ignition to ignition cells by latitude.

It has been demonstrated that highly imbalanced data, such as FF datasets, can adversely affect the performance of ML models [49]. To evaluate the impact of down-sampling, the initial ML modeling test bed was modified for a single RF model test using integer ratios ranging from 1 to 5 in the R function `step_downsample`. The threshold for removing highly correlated features remained fixed at 0.6 for all runs.

Table 4.1: Comparison of the size of the fire cell population (entire province) and the fire cell sample (lower third of provincial mainland).

Boundaries	min lon	max lon	max lat	non-ignition	ignition
Population	−140°	−113°	60°	100,740,455	95,819
Sample	−124°	−115°	53°	25,662,473	65,737
Difference	26°	2°	7°	75,077,982	30,112

The most effective downsampling ratio of non-ignition cells to ignition cells was found to be 1:1 (Table 4.2). This yielded the correct identification of 12,061 true ignitions in the test set, while 2,010 true ignitions were missed.

As anticipated, Recall, also known as Sensitivity, exhibited the same pattern and was lowest when the downsampling ratio was 1:1. There was a consistent increase in Recall and Sensitivity as the downsampling ratio increased. Conversely, Specificity demonstrated a gradual decrease as the downsampling ratio increased. The ROC-AUC value remained stable, resulting in a smooth curve as both Sensitivity increased and Specificity decreased. F1-score proved to be the most suitable metric for assessing the predictive power of the model with regards to non-ignitions (Fig. 4.7). Specificity emerged as the most appropriate metric for evaluating correct and missed ignitions, as shown in Table 4.2 and Fig. 4.6.

Table 4.2: Results of different ignition to non-ignition downsample ratios on RF model performance. (Best result in bold)

downsample non-ign to ign	F1-Score	true non-ign prediction	Specificity	true ign	false non-ign
1:1	0.776	3,591,936	**0.857**	**12061**	**2010**
2:1	0.878	4,435,711	0.708	9967	4104
3:1	0.926	4,896,244	0.545	7669	6402
4:1	0.950	5,135,059	0.448	6304	7767
5:1	**0.976**	**5,417,860**	0.326	4591	9840

4.4 Feature Correlation

The calculation of feature correlation aims to assess the relationships between features in a dataset. It helps identify redundant or highly correlated features, providing insights into the statistical dependence between variables.

To estimate the correlation between pairs of features in our instantiated dataset,

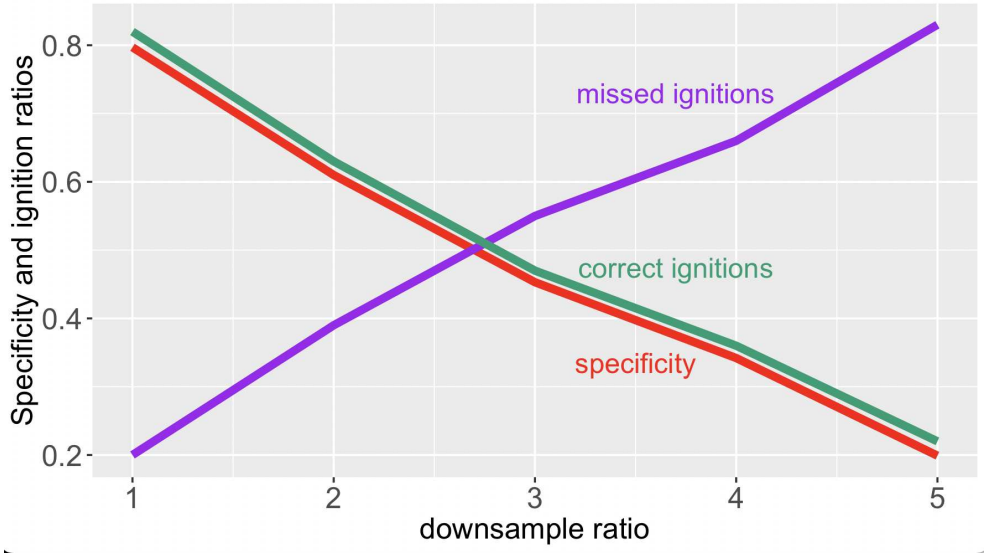


Figure 4.6: Specificity result for different downsample ratios compared to ratios of overall ignitions.

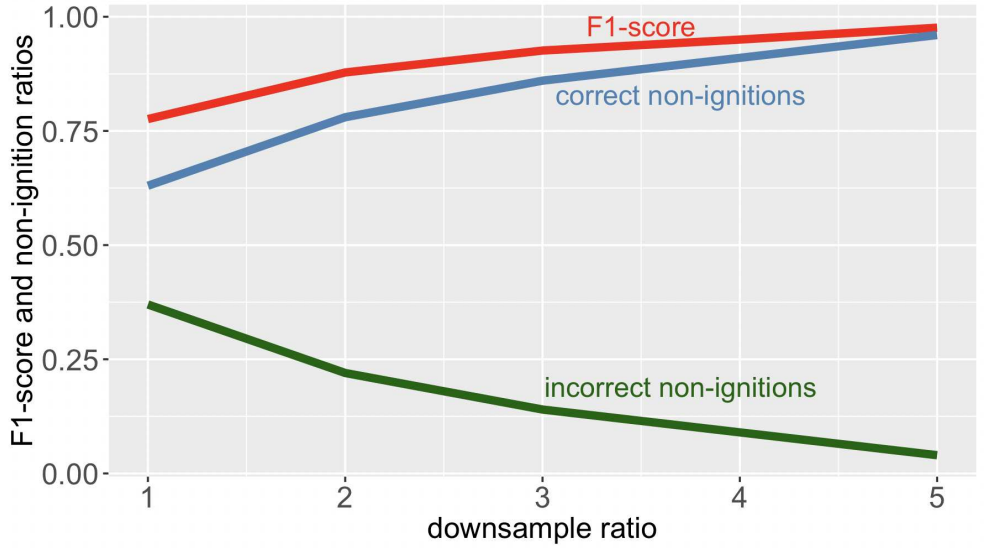


Figure 4.7: F1-score result for different downsample ratios compared to ratios of overall non-ignitions.

we utilized the `Cor` function in the R language. By conducting correlation tests using both Pearson and Spearman Correlation Coefficients, we observed a high correlation of 0.9917707 between Pearson and Spearman results, indicating the adequacy of using Pearson correlation.

Fig. 4.8 presents a correlation matrix heatmap of all features in the novel dataset. Notable observations include the high positive correlation between air temperature

and dew point (d2m), as well as soil temperatures (stl) at various levels. Soil temperatures also exhibit positive correlations among themselves, as do soil moisture content at different levels. Negative correlations can be observed between soil temperatures and soil moisture content, as well as between evaporation and soil temperature.

To assess the impact of feature correlation on the dataset, we applied the `findCorrelation` function in R to remove features based on arbitrarily chosen correlation cutoff points. The function takes the mean absolute correlation of each variable and removes the variable with the largest. Table 4.3 displays the remaining features after adjusting the correlation cutoff. For instance, using a cutoff point of 0.9 resulted in the removal of soil-temperature-level-2 (stl2), skin-temperature (skt), and soil-water-value-level-1 (swvl1) features. The order of feature columns in the novel dataset influenced which features were removed.

To evaluate the effect of altering the number of features used in training, we tested different correlation cutoff points using the RF model. The downsample ratio for all runs was set at 1:1, and the correlation cutoff for feature removal was incremented by 0.1, ranging from 0.6 to 1.0. Table 4.4 presents the results of different correlation cutoff points on the RF model's performance.

Changing the correlation cutoff to remove highly correlated features had minimal impact on the RF model's outcome. There was a marginal improvement in the ratio of correct to missed ignitions, as well as a slight increase in the ROC-AUC value when no features were removed. These results contradict the expected outcome and may be attributed to the use of an RF model. Exploring different models is recommended for future research.

Fig. 4.9 and Fig. 4.10 display the changing rank of features in the feature importance plots for correlation cutoffs of 0.6 and 1.0, respectively. These plots measure the relative importance of a feature across all the trees in an RF model by counting the feature that maximizes the outcome value in each decision tree when used to partition nodes.

When a correlation cutoff of 1.0 is used, where no highly correlated features are removed, soil temperature at various levels and air temperature are the primary influences on the RF model's outcome. However, when a correlation cutoff of 0.6 is applied, resulting in the removal of nine features, the dew point (d2m), soil moisture

evaporation (evabs), surface pressure (sp), and high vegetation coverage (lai_hv) become the key influences. Notably, the dew point remains when using a correlation cutoff of 0.6 due to its column position in the data preceding the other four features. The impact of different column arrangements on the removal of highly correlated features could be explored in future studies.

Latitude does not exhibit high feature importance, potentially due to the sample set's selection from the overall population of ignitions in BC, as shown in Table 4.1. Lowering the upper latitude boundary focused the sample on a region with consistently high numbers of ignitions across all latitudes. If the entire set of available latitudes had been considered, the feature importance of latitude would have been greater due to the uneven distribution of ignitions in BC.

Given that the removal of highly correlated features had minimal impact on the final model outcome, the decision on which key features to use could be based on factors such as ease of deployment, power consumption, and robustness of sensors used for data collection. Additionally, the required precision and size of floating-point numbers for accurate results should be considered when selecting features.

In addition to gauging the effect of correlation on features, we compared the impact of spatio-temporal features to weather features. Comparing the performance of the model created with all features to the model using only weather features demonstrates that temporal and spatial features do have an impact. Missed ignitions totaling 3228 occurred with no temporal or spatial features, as opposed to 2010 when all features are used. The correct ignition prediction to missed ignition prediction ratio for all features is 6.00, as opposed to only weather which gives a ratio of 3.366.

Surprisingly, the year, when combined with weather features, played a significant role in reducing the number of missed ignitions. One possible explanation for this is that the model identified a cyclical pattern over a number of years. The test-train split for this experiment was based on years, with training taking place on years from 1981 to 2010, and testing taking place on years from 2011 to 2020. Future work could try randomizing the years in the two sets.

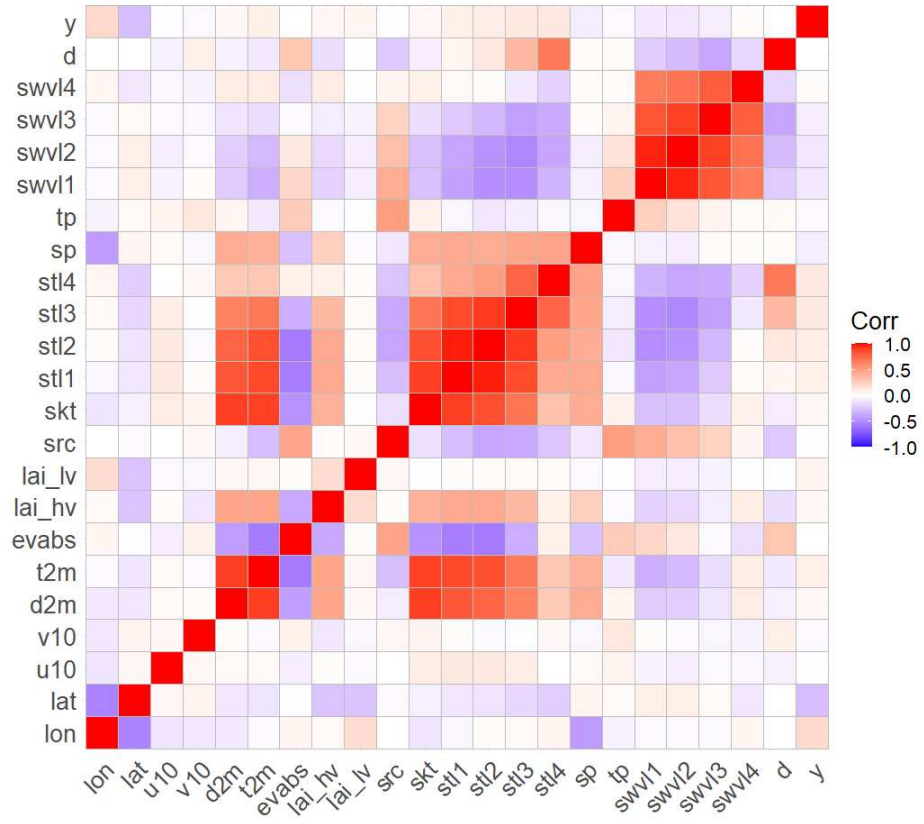


Figure 4.8: Heat map of Pearson Correlation matrix of all features in the novel dataset. (Abbreviation descriptions can be found in Appendix A)

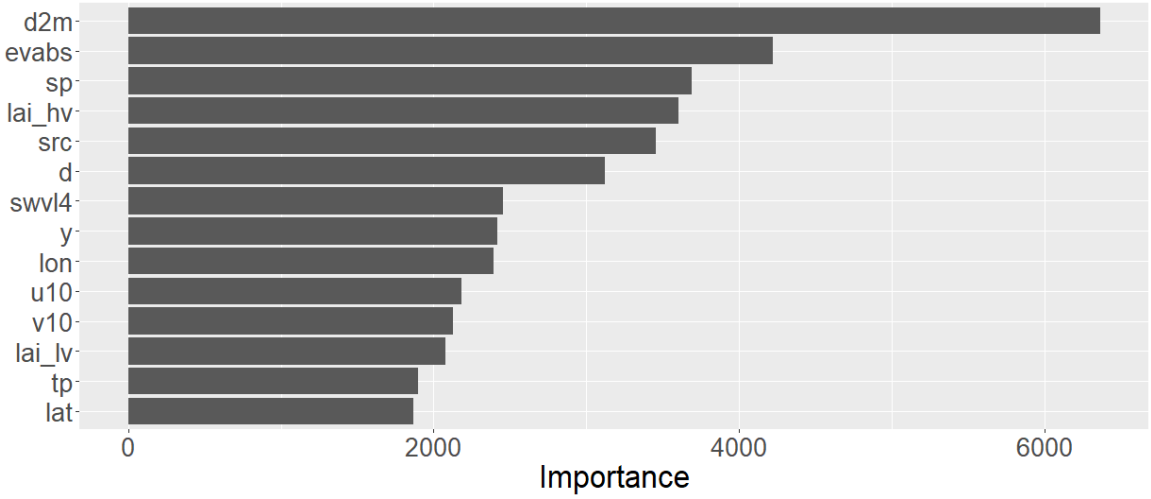


Figure 4.9: Feature importance for a RF model with correlation cutoff of 0.6.

Table 4.3: Features remaining after correlation cutoff adjustment. (Abbreviations found in Appendix A)

cutoff	Features Remaining
0.6	lon, lat, u10, v10, d2m, evabs, lai_hv, lai_lv, src, sp, tp, swvl4, d, y
0.7	lon, lat, u10, v10, d2m, evabs, lai_hv, lai_lv, src, stl4, sp, tp, swvl4, d, y
0.8	lon, lat, u10, v10, d2m, evabs, lai_hv, lai_lv, src, stl4, sp, tp, swvl3, swvl4, d, y
0.9	lon, lat, u10, v10, d2m, t2m, evabs, lai_hv, lai_lv, src, stl1, stl3, stl4, sp, tp, swvl2, swvl3, swvl4, d, y
1.0	lon, lat, u10, v10, d2m, t2m, evabs, lai_hv, lai_lv, src, skt, stl1, stl2, stl3, stl4, sp, tp, swvl1, swvl2, swvl3, swvl4, d, y

Table 4.4: Results of different correlation cutoff points on RF model performance. (Best scores in bold)

correlation cutoff	ROC-AUC	Sens	Spec	true ign	false non-ign
0.6	0.817	0.621	0.857	12061	2010
0.7	0.815	0.615	0.860	12100	1971
0.8	0.817	0.627	0.857	12055	2016
0.9	0.829	0.632	0.870	12245	1826
1.0	**0.830**	**0.638**	**0.869**	**12228**	**1843**

4.5 Model Evaluation Metrics

Our model evaluation primarily relied on Sensitivity, Specificity, and ROC-AUC. Additionally, we used true ignitions (TP) and false non-ignitions (FN) to illustrate the impact of different hyperparameter values, as shown in Tables 4.3, 4.4, and 4.2. We now provide justification for the selection of these metrics over others.

The purpose of evaluation metrics is to measure the alignment between predicted outcomes and observed outcomes. Accuracy is a commonly used measure in ML-based FF prediction, representing the ratio of correct predictions to the total number of predictions. However, accuracy can be misleading when dealing with imbalanced data, such as the instantiated dataset in this study, where the ratio of non-ignition cells to ignition cells is approximately 0.004. A more appropriate approach involves categorizing predictions into TP, FP, TN, and FN. In this study, TP refer to accurate predictions of ignition cells, FP indicate non-ignition cells incorrectly identified

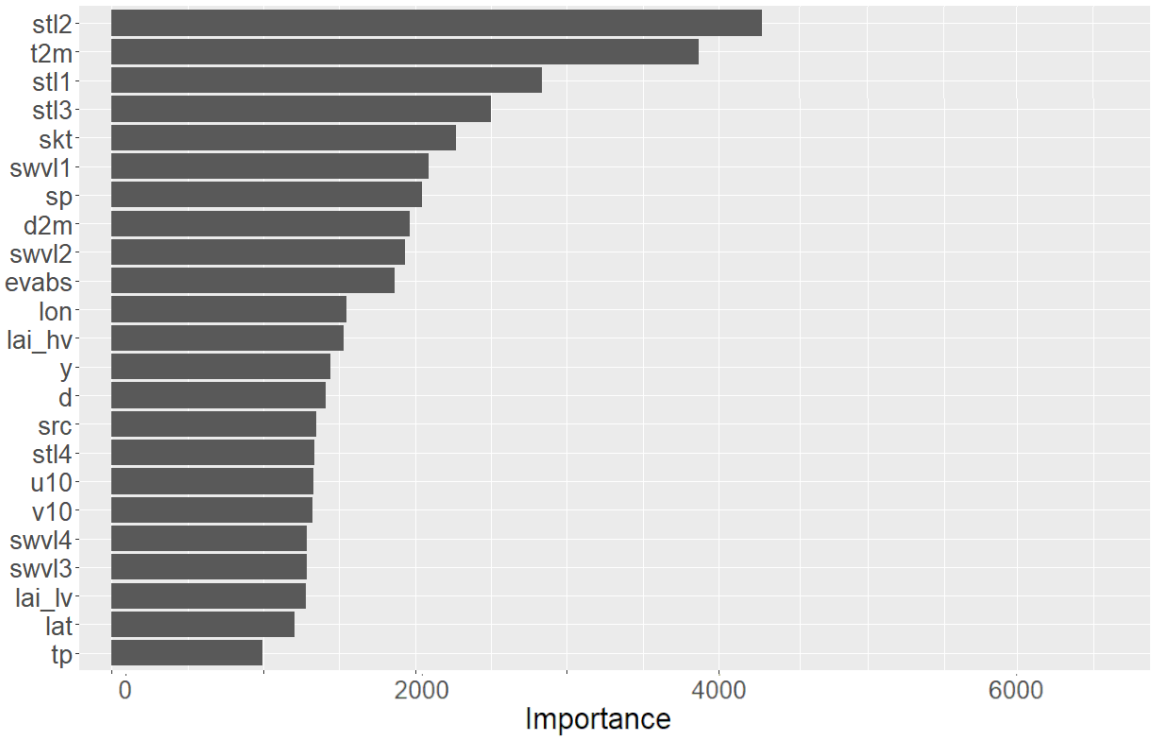


Figure 4.10: Feature importance for a RF model with correlation cutoff of 1.0.

Table 4.5: Sample confusion matrix of observed versus predicted ignitions and non-ignitions.

	predicted non-ignition	predicted ignition
observed non-ignition	true negative	false positive
observed ignition	false negative	true positive

as ignition cells, TN accurately identify non-ignition cells, and FN represent ignition cells misclassified as non-ignition cells (see Table 4.5). An effective FF prediction model strives for a high number of TP and the lowest possible number of FN since misclassifying ignition cells as non-ignition cells poses significant risks to human and forest safety. This true/false classification approach encompasses several metrics, as outlined by Butcher et al. [5], which we employ in this study to evaluate model performance. Sensitivity, also known as the true positive rate, measures the proportion of correctly predicted ignitions. Specificity, also known as the true negative rate, measures the proportion of correctly predicted non-ignitions. Recall is similar to Sensitivity. Recall represents the ratio of correctly predicted ignitions to the sum of correctly predicted ignitions and falsely predicted non-ignitions, providing a measure

of accurately classified ignition instances.

While Sensitivity and Specificity offer insights into model accuracy, they rely on the chosen cutoff value for the prediction probability of ignition or non-ignition. The cutoff value is determined based on the information of interest. For instance, a model emphasizing accurate true predictions would have a different cutoff value than a model aiming to minimize false predictions. To address this issue, we employ the receiver operating characteristic (ROC) curve [5]. The ROC curve is generated by plotting the false positive rate (1-Specificity) against the true positive rate. This curve enables the identification of the optimal cutoff value for the desired prediction information. Additionally, the calculation of the area under the curve (AUC) is valuable. The ROC-AUC takes into account all cutoff values, and a value close to 1 indicates a well-fitting model to the data [5].

4.6 Chapter Summary

This chapter has centered on the evaluation of the dataset's fitness, which was obtained through the instantiation of the data collection framework, through the process of modeling. We have trained ML models using the collected data and assessed their performance using established metrics derived from previous studies.

In the subsequent chapter, we will present a comprehensive framework for FF ignition prediction, employing FL. This FL framework consists of three key components: data, a ML model, and a central server. To evaluate the effectiveness of this framework, we will instantiate it and conduct simulations using a dataset created in Chapter III using the aforementioned framework.

Chapter 5

Federated Learning Framework

In this chapter, we introduce a framework for predicting FF ignitions at specific geographic locations and designated times. Building upon the insights gained from previous chapters, we integrate our knowledge of FF data and ML models to develop a FL system. We instantiate this framework and evaluate its performance through simulations.

5.1 Framework

Our FL framework is comprised of three essential components: data, a ML model, and a CA. The components are represented in Fig. 5.1, with data represented in box 1, the ML model in box 2, prediction in box 3 and the server in box 4. The output prediction of box 3 we define as classifying an ignition or non-ignition leading to a FF in a specific geographic area at a specific time. The local data/local model combination are representative of the IoT WSs in Fig. 1.1 and can be duplicated as required.

The processes within the framework are represented in Fig. 5.2. The data is collected and stored locally (box 1, 2), where it is used to train (box 3), validate, and test a local ML model. The local model weights are then shared with a CA (box 4, 5), which applies the FL averaging of weights (box 6) received from each local model and redistributes them back to each respective local model (box 7, 8, 9). The local models subsequently re-train (box 3) using the averaged weights provided by the CA.

The data primarily consists of weather-related information typically collected by WSs, including temperature, wind speed and direction, and precipitation. The ML model is constrained by the processing and memory capabilities of WSs. A suitable ML model, such as LSTM, capable of capturing relationships in time series data, is employed. The FL framework used in this proposal incorporates simple averaging methods like federated averaging.

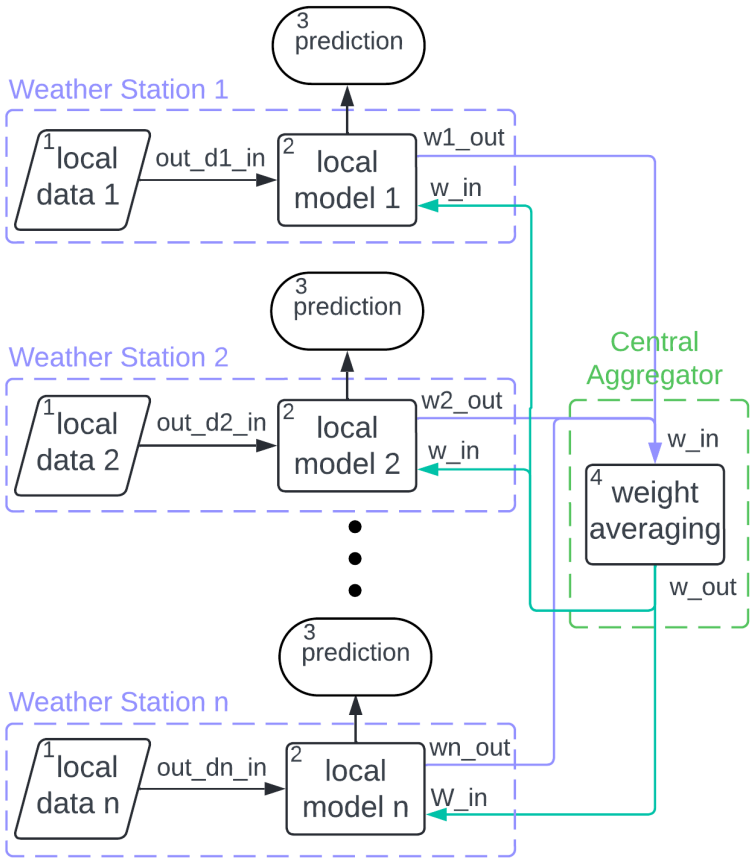


Figure 5.1: A diagrammatic representation of the framework.

5.2 Instantiation Method

To evaluate the performance and effectiveness of integrating FL into platforms like the CWFIS the proposed framework was implemented as a simulation. The objective was to demonstrate the effectiveness of FL when applied to an IoT WS network, as compared to an IoT WS network reporting to a single central server (CS). A comparison of experiment setups is shown in Table 5.4. Key differences between systems are:

- Each local model developed in the FL system used its own simulated IoT WS data whereas the 24 simulated IoT WS data sets were averaged for the CS system.

- Each model in the FL system trained for 3, 6, or 12 epochs per round depending on the experiment whereas the CS system trained for 100 epochs in all

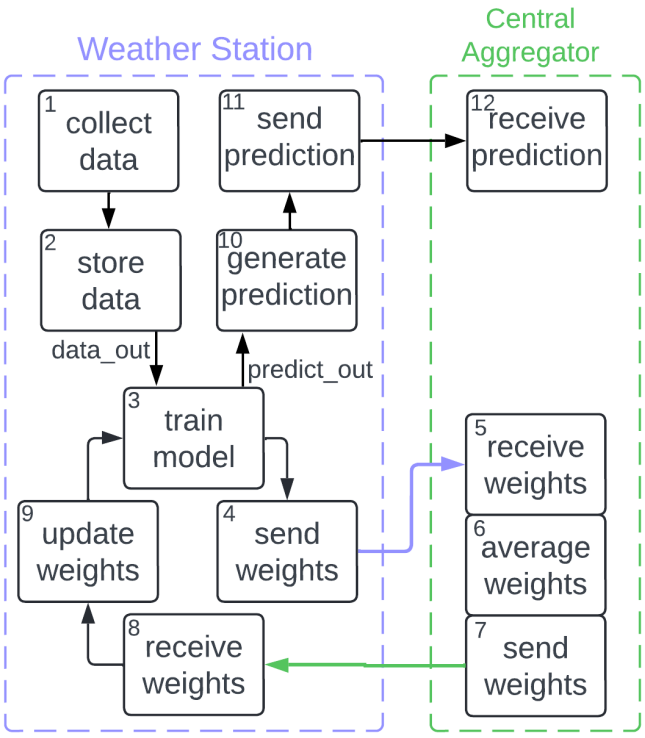


Figure 5.2: Main processes within the framework.

experiments.

- The FL system ran for 3 rounds in experiment 1, 2, and 3, and for 6 rounds in experiment 4. The concept of rounds doesn't apply to the CS system.

Weather data used was reanalysis data obtained from the EMWFC Copernicus project [10]. Motivation and validation for using this type of data can be found in-depth in [42]. The CS system dataset of 5,053,776 rows had 5,008,365 non-ignitions and 45,411 ignitions. This was evenly split among the 24 IoT WSs for the FL system simulation. The ratio of non-ignitions to ignitions varied by station.

5.2.1 Federated Learning Simulation

For the FL system simulation, the fire-prone region surrounding Kelowna, British Columbia was chosen. The region was divided into 24 cells, with 6 cells across and 4 cells in height as represented in Fig. 5.3.

Within each cell 12 data points (blue dots), obtained using reanalysis data, were averaged to produce a simulated IoT WS (orange dots) located at the centre of the cell,

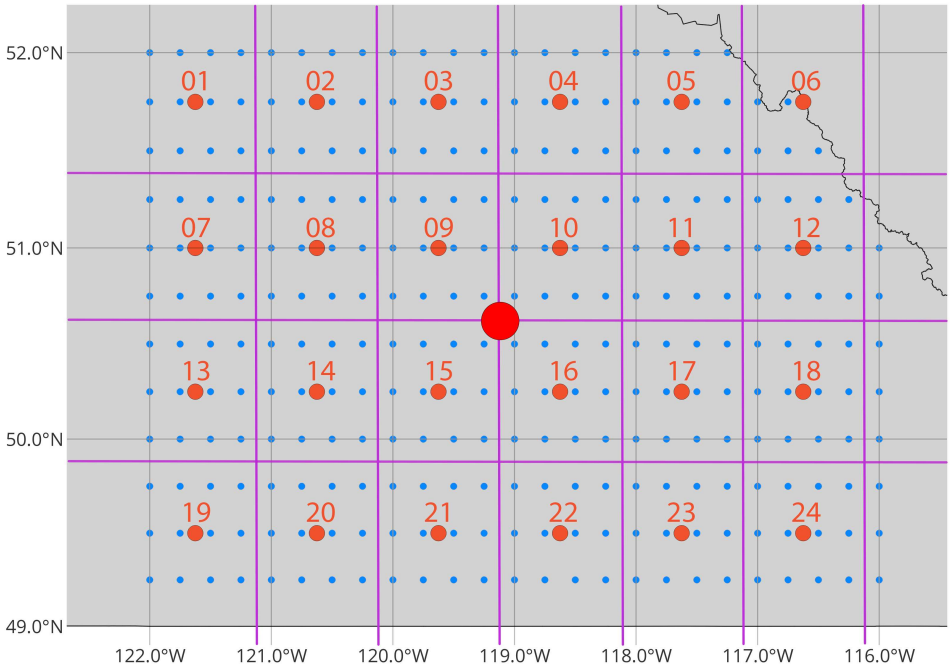


Figure 5.3: Orange = IoT WSs, Blue = Weather Data, Red = CA / CS

Table 5.1: Ignitions in the simulated IoT WS training sets

latitude	Station Ignition Count					
51.75	1,239	1,090	1,036	993	503	142
51.00	1,583	2,919	2,391	1,705	885	852
50.25	1,434	1,964	2,960	1,217	1,090	646
49.50	1,664	1,499	2,914	1,313	1,543	1,103
longitude	121.625	120.625	119.625	118.625	117.625	116.625

Fig. 5.3. The IoT WSs communicate with the CA or CS (single red dot) depending on the system. No communication is assumed between blue and orange dots, as the thesis focuses mainly on the concept demonstration and evaluation of FL.

Simulated IoT WSs had hourly data from 1980 to 2020 containing the weather features: wind direction and speed, dewpoint temperature, air temperature, surface temperature, soil temperature at 4 depths, soil water level at 4 depths, surface pressure, total daily precipitation, soil type, and 2 measures of vegetation coverage. Ignition data from the Government of BC Historical Fire Incidents Location data base was added to the weather features. This resulted in 24 databases, one in each cell, as represented in box 1 of Fig. 5.4. We arbitrarily decided to use 5 days, 120

hours, as the length of our time series. The data format was changed from rows of weather features and ignitions at a specific time and location, Table 5.2, to multivariate time-series data with each row representing a 120 continuous hours of data, Table 5.3 represented in box 2 of Fig. 5.4. The number of ignitions in the training sets of each of the simulated IoT WSs can be seen in Table 5.1. Each database

Table 5.2: Data as table of weather features and ignition class (0, 1).

Longitude	Latitude	Features 1, 2,..., N.	Date	Ign

Table 5.3: Time series data, N is current hour, N-K is K hours before N.

N - K	...	N - 3	N - 2	N - 1	N	Ign

was split into 3 sets. A training set using data from the years 1980-2000, a validation set using data from the years 2001-2010, and a test set using data from the years 2011-2020. This is represented in box 3 of Fig. 5.4. The test and validation data were normalized into a distribution centered around 0 with a standard deviation of 1, box 4 of Fig. 5.4.

Since LSTM models work well with sequential data, and are good at capturing patterns over time such as that seen in weather we chose it as our key model. Its robust implementation in Tensorflow also played a part in our choice. The model, with layers as seen in Fig. 5.5, for each simulated IoT WS was trained on local weather data within each cell to predict FF non-ignition/ignition as seen in box 8 of Fig. 5.4. The model was evaluated by generating predictions 1 day from the last weather input. This was done by offsetting the ignition data 24 hours into the future. To simulate the 24 IoT WSs, one in each cell, maintaining communication with a CA the Flower FL framework [3] was used. For multiple rounds, the simulated CA randomly selected a specified number of simulated IoT WSs and requested their local model weights. The memory available to our simulation machines limited us to 18. The simulated CA then averaged the received weights using the Flower Frameworks built in federated averaging [46]. The averaged weights were sent back to all 24 simulated IoT WSs which updated their model weights. This is represented in boxes 9, 10, 11, 12, 13, and 14 of Fig. 5.4.

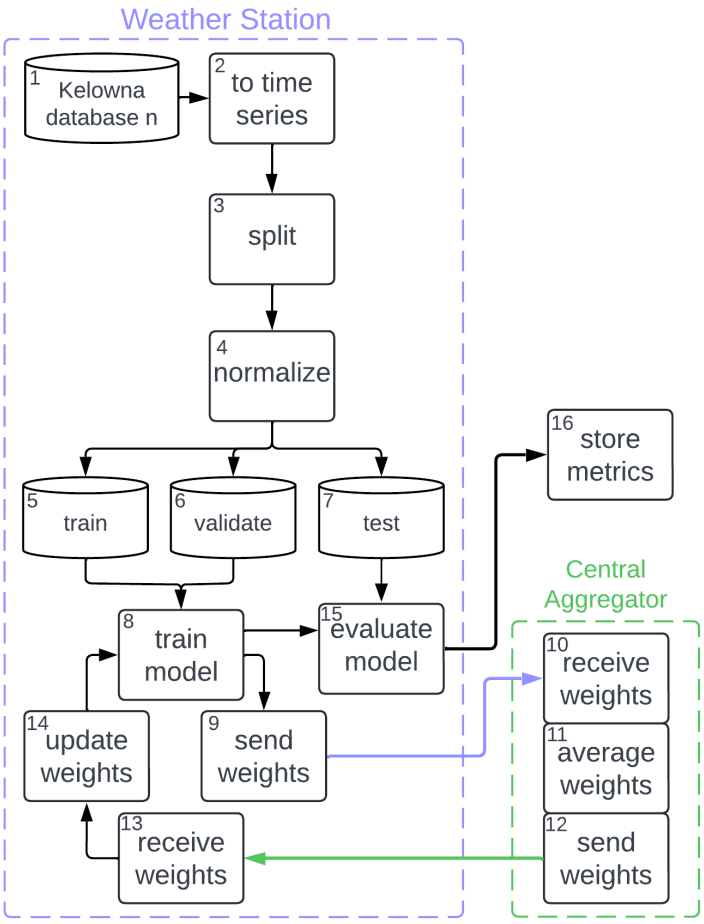


Figure 5.4: Main processes within the instantiation.

When the IoT WSs conducted their next round of local model training, they initiated the process with the new averaged weights, as seen in box 8 of Fig. 5.4. We repeated this cycle of local data collection, prediction, transmission of weights, weight averaging, and local model retraining for 3 rounds. We stopped at 3 as no appreciable improvement was seen.

After a round the local model was evaluated using its test dataset and the results stored in a general database as seen in boxes 9 and 13 of Fig. 5.4. At the end of the simulation the results from all IoT WSs were averaged and compared to the CS system simulation.

Table 5.4: Experiment control and federated setup comparison.

	control	exp 1	exp 2	exp 3	exp 4
system	central	FL	FL	FL	FL
training epochs	100	3	6	12	6
rounds	n/a	3	3	3	6
model	LSTM	LSTM	LSTM	LSTM	LSTM
time series hours	120	120	120	120	120
total stns	n/a	24	24	24	24
training stns/round	n/a	18	18	18	18
eval stns/round	n/a	18	18	18	18

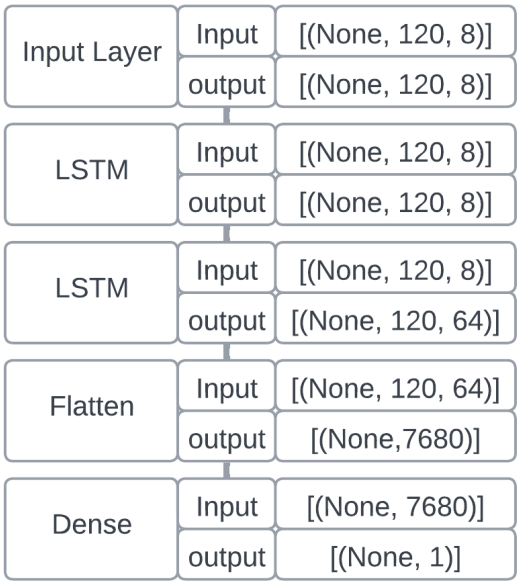


Figure 5.5: The model layers used for all simulations.

5.2.2 CS Simulation

For the CS simulation, the same fire-prone region was chosen and divided in a similar manner to the FL simulation. All of the datasets were combined as seen in boxes 1, 2, 3, and 4 of Fig. 5.6. The rows were converted to time series data. The time series data was split into train, validate and test sets, with the the former two being normalized. This can be seen in boxes 6, 7, 8, 9, and 10 of Fig. 5.6. A LSTM model with the same structure as the FL version was trained and evaluated. The results were stored in the general database with the results of the FL simulation. This last set of processes is represented in boxes 11, 12, and 13 in Fig. 5.6.

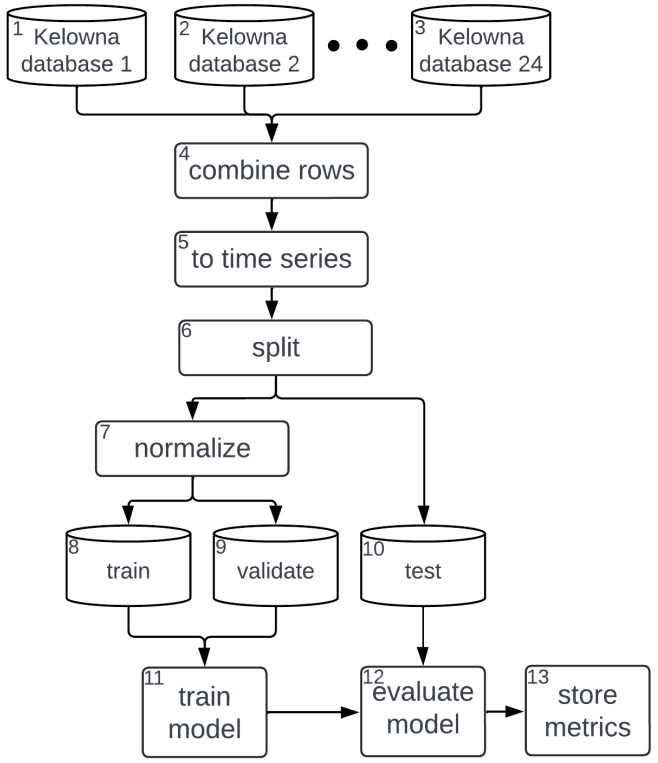


Figure 5.6: The processes for the CS system simulation.

5.3 Instantiation Results

This section presents the results of our simulation, where a FL system used 24 WSs to develop a general model. As a control, we compare the simulation results with those of a CS ML model using the same LSTM layer structure and data. Table 5.5 provides the averaged results of 10 simulation runs, showing the performance of the FL system, which represents an average of 24 IoT WSs. Additionally, Table 5.6

Table 5.5: Averaged results of 10 simulation runs. Federated results are an average of 24 IoT WSs.

experiment	control	exp 1	exp 2	exp 3	exp 4
system type	CS	FL	FL	FL	FL
accuracy	0.7568	0.7453	0.7629	0.7635	0.7603
recall	0.7156	0.6964	0.6892	0.6771	0.6688
ROC-AUC	0.8092	0.8025	0.8076	0.8001	0.7950
sensitivity, spec=0.5	0.9069	0.9014	0.9097	0.9032	0.8985
specificity, sens=0.5	0.8765	0.8699	0.8724	0.8651	0.8596

Table 5.6: A sample of station results from Experiment 2, averaged for 10 simulation runs over 6 epochs/round and 3 training rounds.

station	metrics	round 1	round 2	round 3
1	accuracy	0.7332614437	0.7725694279	0.7627701759
	recall	0.7076401711	0.6565002998	0.6856438696
	ROC-AUC	0.8003950790	0.8002735575	0.8083768368
	sensitivity	0.8812384456	0.8991579413	0.9088108480
	specificity	0.8877710253	0.8703366121	0.8753614545
-	–	–	–	–
8	accuracy	0.7302799463	0.7592176596	0.7686835304
	recall	0.6688471794	0.6857015193	0.6610589847
	ROC-AUC	0.7758552909	0.8028371831	0.7949320897
	sensitivity	0.8586595178	0.8987488846	0.8924262747
	specificity	0.8519785166	0.8705853124	0.8616769984
-	–	–	–	–
16	accuracy	0.7338614000	0.7467972338	0.7666684687
	recall	0.6987206274	0.7209011217	0.6823112170
	ROC-AUC	0.7946205338	0.8150785168	0.8073242505
	sensitivity	0.8827701343	0.9156236947	0.9116604129
	specificity	0.8719882634	0.8804704746	0.8709576428
-	–	–	–	–
24	accuracy	0.7272664383	0.7584168911	0.7622490898
	recall	0.6810897514	0.6855311334	0.6852869391
	ROC-AUC	0.7839112058	0.8038638175	0.8055695295
	sensitivity	0.8754578680	0.9026251435	0.9053723961
	specificity	0.8600174561	0.8725392699	0.8725088909

presents the averaged results for a sample of the 24 stations in Experiment 2 over 6 epochs per round for 3 training rounds.

The results from Experiment 1, 2, 3, and 4 (Table 5.5) indicate that the FL system achieves comparable performance to the CS system. Metrics such as Accuracy, ROC-AUC, Sensitivity, and Specificity show similar values between the two systems. However, the Recall, which measures the proportion of correctly identified ignitions, is slightly higher in the CS system. It is important to note that both systems perform lower than the accuracy of 0.91 achieved by an LSTM model [27] and the recall of 0.97 achieved by a random forest model in previous research [16]. Sensitivity is typically the same as Recall but the results we list are a point on the Sensitivity-Specificity curve obtained when Specificity is 0.5.

Examining the results for individual stations in Experiment 2, (Table 5.6), we observe improvements in most station models. However, Recall shows a decrease as other metrics improve. For instance, station 16 demonstrates improvements in Accuracy, ROC-AUC, and Sensitivity, but a decrease in Recall and Specificity.

In Experiment 3, where the number of training epochs was doubled from 6 to 12, no significant improvements were observed. Instead, a slight reduction in all metrics was observed (Table 5.5). Similarly, in Experiment 4, where the number of rounds

was doubled from 3 to 6, no notable differences were observed compared to the other experiments with only 3 rounds.

We provide communication energy formulas to estimate the data communicated for one round of training in a FL system and a CS system. We consider a round to be the reporting of weather data for a specified time and the subsequent re-training of an existing model using that weather data plus historical data. In the case of a FL system the reporting of model weights back to WSs is considered part of the round.

For the FL system, the formula is defined as the product of the number of IoT WSs (S), the size of the model weights (W_{FL}), the model reporting frequency between each WS and the CA during a round (F_{FL}), and the distance (Dst) from an IoT WS to the CA. We square Dst given that received antenna power density is inversely proportional to the square of the distance between a transmitter-receiver pair. We multiply by a constant of proportionality (C) to resolve units.

$$1)\ E_{\mathrm{FL}} = S_{\mathrm{FL}} \times W_{\mathrm{FL}} \times F_{\mathrm{FL}} \times \sum (Dst^2)_i^{i=S} \times C$$

Similarly, for the CS system, the formula is determined by the number of IoT WSs (S_{CS}), the data size (D_{CS}), the reporting frequency between model training rounds (F_{CS}), and the sum of squared distances from the IoT WSs to the CS:

$$2)\ E_{\mathrm{CS}} = S_{\mathrm{CS}} \times D_{\mathrm{CS}} \times F_{\mathrm{CS}} \times \sum (Dst^2)_i^{i=S} \times C$$

We consider an example with 24 IoT WSs using data (0.094kb/report/station) and weight sizes (20kb/round/station) found in our simulation. For both FL and CS systems WSs record 0.094kb/hour. Each FL system WS uses a single-layer LSTM model with model weights of 20 kb, reporting to and receiving from from the CA once per model training round. Each WS in the CS system reports every hour resulting in a reporting frequency of 168 times (1 week) between model training rounds. Lowest results are shown in bold.

- FL system: $24\mathrm{WS} \times 20kb \times 2 = 960$ kb $\times \sum (Dst^2)_i^{i=S}$
- CS system: $24\mathrm{WS} \times 0.094kb \times 168 = \mathbf{379\ kb} \times \sum (Dst^2)_i^{i=S}$

Increasing the number of model training rounds from 168 to 336 (2 weeks) the data communicated by the FL system remains 960 kb, while the data communicated by the CS system increases to 758 kb.

- FL system: $24\mathrm{WS} \times 20kb \times 2 = 960$ kb $\times \sum (Dst^2)_i^{i=S}$

- CS system: $24\text{WS} \times 0.094kb \times 336 = \mathbf{758\ kb} \times \sum (Dst^2)_i^{i=S}$

Decreasing the number of WS participating in model training within the FL system, results in 18 WS, the data communicated is 720 kb, while the data communicated by the CS system remains at 758 kb, as seen in Fig. 5.7. It is important to note that considering the sum of square of distances between IoT WSs would lead to a greater difference between the FL and CS systems, as the FL system would only sum 18 distances, while the CS system would sum 24.

- FL system: $18\text{WS} \times 20kb \times 2 = \mathbf{720\ kb} \times \sum (Dst^2)_i^{i=S}$
- CS system: $24\text{WS} \times 0.094kb \times 336 = 758\ \text{kb} \times \sum (Dst^2)_i^{i=S}$

These results highlight that initially, the FL system requires greater communication energy due to the larger size of the model weights. However, by decreasing the number of WS involved in training within the FL system or increasing the time between training rounds, Fig. 5.7, the FL system uses less communication energy than the CS system. Additionally, the FL system's ability to evaluate incoming data immediately allows for faster predictions of ignition or non-ignition compared to the CS system, which must first receive transmitted data before making evaluations.

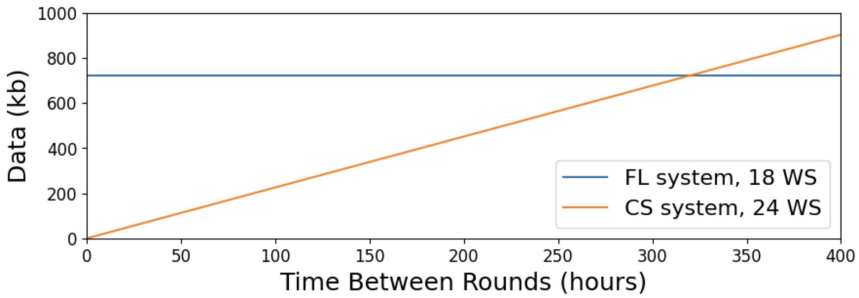


Figure 5.7: Data communicated by FL and CS systems for one round.

5.4 Chapter Summary

In this chapter, we have introduced a FL framework. Building upon the knowledge gained from previous chapters, we integrated our understanding of FF data and ML models to develop a FL system. We instantiated this framework and evaluated its performance through simulations.

The simulations showcased the utilization of IoT data obtained from simulated weather stations in ML models, specifically focusing on the implementation of a LSTM model. This LSTM model was integrated into the FL framework. Altering the number of epochs and rounds in the FL system had minimal impact on the results. While the performance of the FL system was comparable to the CS system, the overall results were lower than what has been observed in previous research [16] [27]. Despite this, the positive impact of integrating FL into platforms like the CWFIS was seen in the spatial accuracy improvements of the FL system over the CS system. Based on the comparable results between systems, we can say that the FL system predicted FF ignitions at a spatial granularity 24 times higher in our simulation. In contrast, the CS system relied on aggregated data from multiple WSs, losing spatial granularity in the process. Additionally, the communication energy (E_{FL}) required in the FL system was initially higher due to the larger size of the model weights (W_{FL}), but if the time between model training is increased or the number of participating WSs (S_{FL}) was reduced, the communication energy (E_{CS}) in the CS system exceeded that of the FL system.

In the final chapter, we will examine the contributions of this thesis, propose potential avenues for future work, and provide a summary of the contributions from this research.

Chapter 6

Conclusion

This thesis makes several significant contributions to the field of forest fire prediction and management.

Firstly, we propose a comprehensive framework for producing high-resolution FF datasets in a standardized .csv format. The framework combines weather reanalysis data with available ignition data and provides provisions for incorporating additional data types such as proximity to human infrastructure and historical lightning data. This framework enables the generation of datasets that represent specific locations and times, including coordinates, feature values, and a binary column indicating the occurrence of an ignition.

Secondly, we instantiate our framework and create three large datasets suitable for simulating IoT weather station data. These datasets vary in temporal and spatial resolutions, incorporating weather features captured daily or hourly over 41 years, along with spatial and FF ignition features. The datasets are formatted in .csv tables, with each row representing a unique day and location.

Thirdly, we evaluate the fitness of these datasets by testing their completeness, correctness, and format validity. By modeling the relationship between weather data and FF ignition data, we demonstrate that the datasets exhibit spatial and temporal coherence, converge in models, adhere to a valid format for standard data processing, and are suitable for use with multiple ML models.

Fourthly, we employ various machine learning models to assess the performance of the datasets in classifying FF ignitions. Using metrics such as F1-score, ROC-AUC, Sensitivity, and Specificity, we evaluate the models' ability to correctly identify ignitions within a given geographic area and time interval. The tested models include Bag Tree, Boost Tree, Decision Tree, Logistic Regression, Multilayer Perceptron, Nearest Neighbor, Random Forest, and Support Vector Machine.

We demonstrate the practical value of integrating weather reanalysis data with

FF ignition data by achieving ROC-AUC and F1-score values above 0.8. This highlights the effectiveness of combining these datasets and underscores the potential for enhanced FF prediction accuracy.

FInally, we introduce a novel framework that combines FL with simulated data collection from IoT devices, specifically WSs. Through the instantiation of this framework as a simulation using a database from the Kelowna region, we evaluate its effectiveness and applicability within a localized and representative context. The simulation demonstrated that utilizing data gathered through IoT WSs in ML models, specifically LSTM, can yield comparable prediction outcomes to a CS model trained on aggregated data. In addition, the spatial granularity of the FF ignition classification predictions is higher in the FL system by the nature of each WS maintaining its own model.

Furthermore, the FL system demonstrated advantages in terms of communication energy and latency. By distributing the model training process among individual WSs, the FL system reduces the amount of data transferred, resulting in lower communication energy. Additionally, each WS can evaluate data promptly, leading to reduced latency compared to the CS system.

While our thesis presents valuable frameworks and methodologies for forest fire prediction and management, there are potential limitations and areas that warrant further investigation.

Firstly, one potential concern lies in the generalizability of the two frameworks we have developed. These frameworks have been instantiated and evaluated in a specific geographic region, which may limit their applicability to other regions with different environmental and fire conditions.

Secondly, the choice of machine learning models utilized in the FL simulation could be subject to further investigation. While we have employed various models, it is essential to explore other models that may be better suited to the specific problem of forest fire prediction. Exploring alternative models could potentially enhance the accuracy and efficiency of the prediction system.

Additionally, expanding the data collection methods beyond IoT weather stations could be beneficial. While weather stations provide valuable data, incorporating other data collection techniques, such as unmanned aerial vehicles (UAVs), could offer

broader coverage, especially in large remote regions where access may be limited.

In summary, this thesis contributes a comprehensive framework for generating FF datasets, instantiated datasets suitable for machine learning, an evaluation of dataset fitness and model performance, and a novel framework integrating FL and IoT technologies.

Our federated learning framework affirms the potential of ML, FL, and IoT in enhancing FF prediction capabilities within platforms like the CWFIS. The study contributes to the understanding and mitigation of the growing threat posed by FF. By exploring technologies and frameworks such as FL and IoT, fire management agencies can leverage the power of historical and real-time data as well as ML to make informed decisions and effectively respond to FF risks.

Bibliography

[1] Rdocumentation, terra package. accessed: 13.05.2023.

[2] Faroudja Abid, Nouma Izeboudjen, and Fatiha Louiz. Hardware/software development of a machine learning based forest fires prediction system. In *2021 International Conference on Information Systems and Advanced Technologies (ICISAT)*, pages 1–5, 2021.

[3] Daniel J. Beutel, Taner Topal, Akhil Mathur, Xinchi Qiu, Javier Fernandez-Marques, Yan Gao, Lorenzo Sani, Kwing Hei Li, Titouan Parcollet, Pedro Porto Buarque de Gusmão, and Nicholas D. Lane. Flower: A friendly federated learning research framework, 2022.

[4] Tanuj Bhatt and Arashdeep Kaur. Automated forest fire prediction systems: A comprehensive review. In *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pages 1–5, 2021.

[5] Brandon Butcher and Brian J. Smith. Feature engineering and selection: A practical approach for predictive models. *The American Statistician*, 74(3):308–309, 2020.

[6] Natural Resources Canada. The state of canada's forests: Annual report 2022. 2022.

[7] Statistics Canada. 2021 census boundary files. `https://www12.statcan.gc.ca/census-recensement/2021/geo/sip-pis/boundary-limites/index2021-eng.cfm?year=21`. (accessed: 27.09.2022).

[8] Pacific Climate Impacts Consortium. Meeting the need for practical climate information — pacific climate impacts consortium. `https://www.pacificclimate.org`. (accessed: 27.09.2022).

[9] Pacific Climate Impacts Consortium. Terms of use. `https://www.pacificclimate.org/terms-of-use/`. (accessed: 27.09.2022).

[10] CopernicusECMWF. Climate reanalysis — copernicus. `https://climate.copernicus.eu/climate-reanalysis`. (accessed: 27.09.2022).

[11] CopernicusECMWF. How to acknowledge and cite a climate data store (cds) catalogue entry and the data published as part of it — copernicus. `https://confluence.ecmwf.int/display/CKB/How+to+acknowledge+and+cite+a+Climate+Data+Store+%28CDS%29+catalogue+entry+and+the+data+published+as+part+of+it`. (accessed: 07.11.2022).

[12] Gao Demin, Lin Haifeng, Jiang Anna, and Wu Guoxin. A forest fire prediction system based on rechargeable wireless sensor networks. In *2014 4th IEEE International Conference on Network Infrastructure and Digital Content*, pages 405–408, 2014.

[13] Hay Devona. Fire incident locations - historical. `https://www2.gov.bc.ca/gov/content/data/open-data/open-government-licence-bc`, 2022. Accessed: 15.07.2022.

[14] Osama Elsarrar, Marjorie Darrah, and Richard Devine. Analysis of forest fire data using neural network rule extraction with human understandable rules. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 1917–19176, 2019.

[15] Jennifer Gabrys. Smart forests and data practices: From the internet of trees to planetary governance. *Big Data & Society*, 7(1):2053951720904871, 2020.

[16] Stella Girtsou, Alexis Apostolakis, Giorgos Giannopoulos, and Charalampos Kontoes. A machine learning methodology for next day wildfire prediction. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 8487–8490, 2021.

[17] Nurdeka Hidayanto, Adhi Harmoko Saputro, and Danang Eko Nuryanto. Peatland data fusion for forest fire susceptibility prediction using machine learning. In *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pages 544–549, 2021.

[18] Fantine Huot, R. Lily Hu, Nita Goyal, Tharun Sankar, Matthias Ihme, and Yi-Fan Chen. Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022.

[19] Tanqiu Jiang, Sidhant K. Bendre, Hanjia Lyu, and Jiebo Luo. From static to dynamic prediction: Wildfire risk assessment based on multiple environmental factors. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 4877–4886, 2021.

[20] Harkiran Kaur and Sandeep Kumar Sood. Energy-efficient iot-fog-cloud architectural paradigm for real-time wildfire prediction and forecasting. *IEEE Systems Journal*, 14(2):2003–2011, 2020.

[21] M. Kuhn and J. Silge. *Tidy Modeling with R.* O'Reilly Media, 2022.

[22] Max Kuhn and Kjell Johnson. *Applied predictive modeling.* Springer, New York, NY, 2019.

[23] Nihal Kumar and Aditya Kumar. Australian bushfire detection using machine learning and neural networks. In *2020 7th International Conference on Smart Structures and Systems (ICSSS)*, pages 1–7, 2020.

[24] Ryan Laube and Howard J. Hamilton. Wildfire occurrence prediction using time series classification: A comparative study. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 4178–4182, 2021.

[25] B. D. Lawson and O. B. Armitage. Weather guide for the canadian forest fire danger rating system. 2008.

[26] Bryan S. Lee. The canadian wildland fire information system. In *Proceedings of the Ninth Annual Symposium on Geographic Systems*, Vancouver, Canada, 1995. Publisher.

[27] Hao Liang, Meng Zhang, and Hailan Wang. A neural network model for wildfire scale prediction using meteorological factors. *IEEE Access*, 7:176746–176755, 2019.

[28] Ashima Malik, Nasrajan Jalin, Shalu Rani, Priyanka Singhal, Supriya Jain, and Jerry Gao. Wildfire risk prediction and detection using machine learning in san diego, california. In *2021 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/IOP/SCI)*, pages 622–629, 2021.

[29] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data, 2023.

[30] Slobodan Milanović, Nenad Marković, Dragan Pamučar, Ljubomir Gigović, Pavle Kostić, and Sladjan D. Milanović. Forest fire probability mapping in eastern serbia: Logistic regression versus random forest method. *Forests*, 12(1), 2021.

[31] K Venkata Murali Mohan, Aravapalli Rama Satish, K Mallikharjuna Rao, Rokesh Kumar Yarava, and G Charles Babu. Leveraging machine learning to predict wild fires. In *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, pages 1393–1400, 2021.

[32] A. Mutakabbir, C-H Lung, S. A. Ajila, M. Zaman, K. Naik, R. Purcell, and S. Sampalli. Spatio-temporal agnostic deep learning modeling of forest fire prediction using weather data. In *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*, Torino, Italy, 2023.

[33] Abdul Mutakabbir. Mutli-modality federate learning with multi-source data for forest fire prediction, 05 2023.

[34] J Muñoz Sabater. Era5-land hourly data from 1981 to present., 2021. Accessed: 15.07.2022.

[35] NASA. Fire information for resource management system (firms). https://www.earthdata.nasa.gov/learn/find-data/near-real-time/firms. (accessed Jan. 10, 2023).

[36] Saily Natekar, Shivani Patil, Aishwarya Nair, and Sukanya Roychowdhury. Forest fire prediction using lstm. In *2021 2nd International Conference for Emerging Technology (INCET)*, pages 1–5, 2021.

[37] Tuan Nguyen-Anh and Quan Le-Trung. Prediction of forest fire risk to trigger iots reconfiguration action. In *2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, pages 19–24, 2020.

[38] Mohammad Reza Nosouhi, Keshav Sood, Neeraj Kumar, Tricia Wevill, and Chandra Thapa. Bushfire risk detection using internet of things: An application scenario. *IEEE Internet of Things Journal*, 9(7):5266–5274, 2022.

[39] Government of Alberta. Wildfire maps and data. https://www.alberta.ca/wildfire-maps-and-data.aspx#jumplinks-2. (accessed: 04.11.2023).

[40] Government of BC. Open government licence - british columbia. https://www2.gov.bc.ca/gov/content/data/open-data/open-government-licence-bc. (accessed: 27.09.2022).

[41] T Preeti, Suvarna Kanakaraddi, Aishwarya Beelagi, Sumalata Malagi, and Aishwarya Sudi. Forest fire prediction using machine learning techniques. In *2021 International Conference on Intelligent Technologies (CONIT)*, pages 1–6, 2021.

[42] R. Purcell, K. Naik, S. Sampalli, C. Lung, M. Zaman, A. Mutakabbir, P. Kaur, and F. Tavakoli. A framework for creating forest fire ignition prediction datasets. Under review, 2023.

[43] Pradeep Kumar Singh and Amit Sharma. An insight to forest fire detection techniques using wireless sensor networks. In *2017 4th International Conference on Signal Processing, Computing and Control (ISPCC)*, pages 647–653, 2017.

[44] Statology. How to perform logistic regression in r (step-by-step). https://www.statology.org/logistic-regression-in-r/. (accessed: 04.01.2023).

[45] U.S. Geological Survey. Combined wildfire datasets for the united states and certain territories, 1878-2019. https://www.sciencebase.gov/catalog/item/5ee13de982ce3bd58d7be7e7. accessed: 15.04.2023.

[46] Flower Development Team. Flower: Intro to fl pytorch. https://flower.dev/docs/tutorial/Flower-1-Intro-to-FL-PyTorch.html. accessed: 15.04.2023.

[47] Ranak Thakkar, Varad Abhyankar, Polaka Divya Reddy, and Surya Prakash. Environmental fire hazard detection and prediction using random forest algorithm. In *2022 International Conference for Advancement in Technology (ICONAT)*, pages 1–4, 2022.

[48] Stefania Tomasiello and Muhammad Uzair. Some remarks on anfis for forest fires prediction. In *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–5, 2021.

[49] Jason Van Hulse, Taghi M. Khoshgoftaar, and Amri Napolitano. Experimental perspectives on learning from imbalanced data. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, page 935–942, New York, NY, USA, 2007. Association for Computing Machinery.

[50] Jin Zou, Ning Lu, Hou Jiang, Jun Qin, Ling Yao, Ying Xin, and Fenzhen Su. Performance of air temperature from era5-land reanalysis in coastal urban agglomeration of southeast china. *Science of The Total Environment*, 828:154459, 2022.

Appendix A

Lower BC Dataset Feature Descriptions

Table A.1: Description of features in the NetCDF downloaded from ECMWF.

Feature	Unit	Description
lon	degrees	Upper left hand corner longitude of a cell.
lat	degrees	Upper left hand corner latitude of a cell.
u10	m/s	Eastward component of wind at 10m above the earth's surface.
v10	m/s	Northward component of wind at 10m above the earth's surface.
d2m	Kelvin (K)	The dewpoint temperature at 2m above the earth's surface. It is a measure of humidity.
t2m	Kelvin (K)	Air temperature at 2m above the earth's surface.
evabs	m of water	The amount of water evaporation from bare soil.
lai_hv	m^2/m^2	One-half of the total green leaf area per unit horizontal ground surface area for high vegetation.
lai_lv	m^2/m^2	One-half of the total green leaf area per unit horizontal ground surface area for low vegetation.
src	m of water	The skin reservoir content is the amount of water in the vegetation canopy and/or in a thin layer on the soil.
skt	Kelvin (K)	Skin temperature is the temperature at the earth's surface.
stl1	Kelvin (K)	Temperature of the soil in layer 1 (0 -7 cm) of the ECMWF Integrated Forecasting System.
stl2	Kelvin (K)	Temperature of the soil in layer 2 (7 -28 cm) of the ECMWF Integrated Forecasting System.
stl3	Kelvin (K)	Temperature of the soil in layer 3 (28 -100 cm) of the ECMWF Integrated Forecasting System.
stl4	Kelvin (K)	Temperature of the soil in layer 1 (100 -289 cm) of the ECMWF Integrated Forecasting System.
sp	Pa	Surface pressure (force per unit area)is the atmospheric pressure at the earth's surface.
tp	m	Total precipitation.
swvl1	m^3/m^3	Volume of water in soil layer 1 (0 -7 cm) of the ECMWF Integrated Forecasting System.
swvl2	m^3/m^3	Volume of water in soil layer 1 (7 -28 cm) of the ECMWF Integrated Forecasting System.
swvl3	m^3/m^3	Volume of water in soil layer 1 (28-100 cm) of the ECMWF Integrated Forecasting System.
swvl4	m^3/m^3	Volume of water in soil layer 1 (100 -289 cm) of the ECMWF Integrated Forecasting System.
d	day	The day of the year for the given year.
y	year	The year.
ignition	boolean	1 if any ignition occured in the cell on the day, otherwise 0.
ttl_ign	integer	Total ignitions in the cell on the day.
person	integer	Total ignitions started by persons on the day.
ligtning	integer	Total ignitions started by lightning of the day.
unknown	integer	Total ignitions with unknown origin on the day.
fire_id	character	All fire ID's for the cell on the day.

Features in output order. For more information see Muñoz [34]

Appendix B

Lower BC Dataset Instantiation Example

Table B.1: Column headings and 2 rows of BC dataset.

gridID	lon	lat	u10	v10	d2m
1	-139	60	-0.70801	-0.75784	267.618
2	-138.9	60	-0.61414	-0.81431	266.942

t2m	evabs	lai_hv	lai_lv	src	skt
268.8362	0	4.44E-16	0	0.00019	268.5468
268.2463	0	4.44E-16	0	0.000193	268.0623

stl1	stl2	stl3	stl4	sp	tp
263.3696	263.2085	262.6628	262.1564	84669.66	0.011862
262.676	262.5093	261.9791	261.3105	83623.83	0.011453

swvl1	swvl2	swvl3	swvl4	d	y
0.586042	0.583066	0.179998	0.342602	122	1980
0.370822	0.376846	0.188451	0.360737	122	1980

ignition	ttl_ign	fire_id	person	lightning	unknown
0	0	0	0	0	0
0	0	0	0	0	0

Appendix C

Kelowna Dataset Features

Table C.1: Description of features for Kelowna Dataset.

Feature	Unit	Description
date	seconds	Time passed since 1900 in seconds.
lon	degrees	Upper left hand corner longitude of a cell.
lat	degrees	Upper left hand corner latitude of a cell.
u10	m/s	Eastward component of wind at 10m above the earth's surface.
v10	m/s	Northward component of wind at 10m above the earth's surface.
d2m	Kelvin (K)	The dewpoint temperature at 2m above the earth's surface. It is a measure of humidity.
t2m	Kelvin (K)	Air temperature at 2m above the earth's surface.
e	m of water	The amount of water that has evaporated from the earth's surface.
cvh	m^2/m^2	One-half of the total green leaf area per unit horizontal ground surface area for high vegetation.
cvl	m^2/m^2	One-half of the total green leaf area per unit horizontal ground surface area for low vegetation.
slt	integer	The soil type based on texture. 1: Coarse, 2: Medium, 3: Medium fine, 4: Fine, 5: Very fine, 6: Organic, 7: Tropical organic. Zero is a non-land point.
skt	Kelvin (K)	Skin temperature is the temperature at the earth's surface.
stl1	Kelvin (K)	Temperature of the soil in layer 1 (0 -7 cm) of the ECMWF Integrated Forecasting System.
stl2	Kelvin (K)	Temperature of the soil in layer 2 (7 -28 cm) of the ECMWF Integrated Forecasting System.
stl3	Kelvin (K)	Temperature of the soil in layer 3 (28 -100 cm) of the ECMWF Integrated Forecasting System.
stl4	Kelvin (K)	Temperature of the soil in layer 1 (100 -289 cm) of the ECMWF Integrated Forecasting System.
sp	Pa	Surface pressure (force per unit area)is the atmospheric pressure at the earth's surface.
tp	m	Total precipitation.
swvl1	m^3/m^3	Volume of water in soil layer 1 (0 -7 cm) of the ECMWF Integrated Forecasting System.
swvl2	m^3/m^3	Volume of water in soil layer 1 (7 -28 cm) of the ECMWF Integrated Forecasting System.
swvl3	m^3/m^3	Volume of water in soil layer 1 (28-100 cm) of the ECMWF Integrated Forecasting System.
swvl4	m^3/m^3	Volume of water in soil layer 1 (100 -289 cm) of the ECMWF Integrated Forecasting System.
ignition	boolean	1 if ignition occurred in the cell on the day.
month	integer	The month of the year for the given year.
day	integer	The day of the month.
hour	integer	The hour for the given day.
year	integer	The year.
doy	integer	The day of the year for the given year.
year	year	The year.

Appendix D

Kelowna Dataset Summary

Table D.1: Summary of Kelowna features values.

features	min	max	range	mean	sd
date	3.23E+08	1.60E+09	1.28E+09	9.64E+08	3.73E+08
lon	-122	-116	6	-119	1.802776
lat	49	52	3	50.5	0.935414
u10	-7.05325	8.853455	15.90671	0.346685	1.144646
v10	-7.30391	9.983047	17.28696	0.400354	1.153399
d2m	234.8329	296.2841	61.45123	275.4998	5.214091
t2m	238.0774	310.1319	72.05443	282.0397	7.303597
e	-0.00073	8.10E-05	0.000816	-7.47E-05	9.62E-05
cvh	0.156647	0.999878	0.843231	0.878662	0.15677
cvl	3.06E-05	0.622479	0.622448	0.10449	0.131285
skt	233.5343	315.5637	82.02937	280.8523	8.187633
stl1	247.226	307.8882	60.6622	280.752	6.315398
stl2	256.7656	299.6206	42.85498	280.4581	5.40597
stl3	270.6714	293.3479	22.67651	279.7688	4.591183
stl4	272.2055	288.9502	16.74467	278.4565	3.946321
slt	1	2	1	1.859107	0.347912
sp	75566.63	96598.5	21031.87	86133.18	2971.832
tp	-8.67E-19	0.012767	0.012767	0.000105	0.000292
swvl1	0.030548	0.457254	0.426705	0.316099	0.069188
swvl2	0.075025	0.453682	0.378657	0.315822	0.063187
swvl3	0.100568	0.453829	0.353261	0.315199	0.055933
swvl4	0.211334	0.437897	0.226563	0.351139	0.040686
month	4	10	6	7.009346	1.99764
day	1	31	30	15.78972	8.827405
hour	0	23	23	11.5	6.922187
ignition	0	1	1	0.000759	0.027533
year	1980	2020	40	2000	11.83216
doy	91	305	214	197.7683	61.77739

Appendix E

High Prairie Dataset Features

Table E.1: Description of features for High Prairie Dataset.

Feature	Unit	Description
date	seconds	Time passed since 1900 in seconds.
lon	degrees	Upper left hand corner longitude of a cell.
lat	degrees	Upper left hand corner latitude of a cell.
u10	m/s	Eastward component of wind at 10m above the earth's surface.
v10	m/s	Northward component of wind at 10m above the earth's surface.
d2m	Kelvin (K)	The dewpoint temperature at 2m above the earth's surface. It is a measure of humidity.
t2m	Kelvin (K)	Air temperature at 2m above the earth's surface.
e	m of water	The amount of water that has evaporated from the earth's surface.
cvh	m^2/m^2	One-half of the total green leaf area per unit horizontal ground surface area for high vegetation.
cvl	m^2/m^2	One-half of the total green leaf area per unit horizontal ground surface area for low vegetation.
slt	integer	The soil type based on texture. 1: Coarse, 2: Medium, 3: Medium fine, 4: Fine, 5: Very fine, 6: Organic, 7: Tropical organic. Zero is a non-land point.
skt	Kelvin (K)	Skin temperature is the temperature at the earth's surface.
stl1	Kelvin (K)	Temperature of the soil in layer 1 (0 -7 cm) of the ECMWF Integrated Forecasting System.
stl2	Kelvin (K)	Temperature of the soil in layer 2 (7 -28 cm) of the ECMWF Integrated Forecasting System.
stl3	Kelvin (K)	Temperature of the soil in layer 3 (28 -100 cm) of the ECMWF Integrated Forecasting System.
stl4	Kelvin (K)	Temperature of the soil in layer 1 (100 -289 cm) of the ECMWF Integrated Forecasting System.
sp	Pa	Surface pressure (force per unit area)is the atmospheric pressure at the earth's surface.
tp	m	Total precipitation.
swvl1	m^3/m^3	Volume of water in soil layer 1 (0 -7 cm) of the ECMWF Integrated Forecasting System.
swvl2	m^3/m^3	Volume of water in soil layer 1 (7 -28 cm) of the ECMWF Integrated Forecasting System.
swvl3	m^3/m^3	Volume of water in soil layer 1 (28-100 cm) of the ECMWF Integrated Forecasting System.
swvl4	m^3/m^3	Volume of water in soil layer 1 (100 -289 cm) of the ECMWF Integrated Forecasting System.
ignition	boolean	1 if ignition occurred in the cell on the day.
month	integer	The month of the year for the given year.
day	integer	The day of the month.
hour	integer	The hour for the given day.
year	integer	The year.
doy	integer	The day of the year for the given year.
year	year	The year.

Appendix F

High Prairie Dataset Summary

Table F.1: Summary of High Prairie features values.

features	min	max	range	mean	sd
lon	-117.5	-114.5	3	-116	0.935414
lat	50.5	57.5	7	54	2.09165
u10	-10.1597	14.99162	25.15137	0.757982	2.245168
v10	-13.5863	11.39349	24.9798	0.296311	1.826674
d2m	233.2032	297.2014	63.99825	275.7906	6.69871
t2m	236.9807	309.4077	72.427	282.2494	7.799168
e	-0.00099	9.46E-05	0.001087	-8.70E-05	0.000107
cvh	0	1	1	0.867875	0.212209
cvl	0	0.752797	0.752797	0.111972	0.187168
skt	231.6383	312.5334	80.89516	282.0122	8.284807
stl1	251.701	305.6262	53.92522	281.7348	6.587747
stl2	260.2542	300.6533	40.39917	281.2484	5.725052
stl3	265.9739	299.9766	34.00269	279.7083	5.119141
stl4	271.281	299.2822	28.00122	277.7643	4.092277
slt	0	6	6	2.756003	1.461605
sp	73894.63	99497.06	25602.44	90064.24	5074.583
tp	-8.67E-19	0.019022	0.019022	9.32E-05	0.00032
swvl1	0.016958	0.777129	0.760171	0.368325	0.10626
swvl2	0.02402	0.769423	0.745403	0.364521	0.103622
swvl3	0.030157	0.765159	0.735002	0.358088	0.100677
swvl4	0.033157	0.749084	0.715927	0.368113	0.089542
month	4	10	6	7.009346	1.99764
day	1	31	30	15.78972	8.827405
hour	0	23	23	11.5	6.922187
date	2.60E+08	1.54E+09	1.28E+09	9.01E+08	3.73E+08

Appendix G

Additional Federated Learning Simulation Results

Table G.1: Experiment 2 averaged results for 10 simulation runs of stations 1-7 over 6 epochs/round and 3 training rounds.

station	metrics	round 1	round 2	round 3
1	accuracy	0.7332614437	0.7725694279	0.7627701759
	recall	0.7076401711	0.6565002998	0.6856438696
	ROC-AUC	0.800395079	0.8002735575	0.8083768368
	sensitivity	0.8812384456	0.8991579413	0.908810848
	specificity	0.8877710253	0.8703366121	0.8753614545
2	accuracy	0.7310531395	0.7612825572	0.758407943
	recall	0.7047800081	0.6892268717	0.6979562789
	ROC-AUC	0.7986696448	0.8085339963	0.8106391057
	sensitivity	0.88059026	0.9055766761	0.9109632447
	specificity	0.8854282073	0.8782616854	0.8785534054
3	accuracy	0.7330729195	0.7665400604	0.7572401389
	recall	0.6799926673	0.6869918803	0.6975288838
	ROC-AUC	0.7876956207	0.8129883409	0.810118027
	sensitivity	0.8728222932	0.9144201974	0.9114248902
	specificity	0.8690363254	0.8788161278	0.8758484945
4	accuracy	0.734152548	0.7623437524	0.7599593912
	recall	0.7137335539	0.6928947389	0.7057330779
	ROC-AUC	0.804886654	0.8123638391	0.817211645
	sensitivity	0.8992598727	0.917368418	0.9218985013
	specificity	0.8796337247	0.8751379073	0.8784800087
5	accuracy	0.7304916581	0.7579173843	0.7575701773
	recall	0.6963113811	0.6894470851	0.7089839379
	ROC-AUC	0.794596937	0.8088807133	0.8204431633
	sensitivity	0.8842121098	0.9132533802	0.9218148887
	specificity	0.8751309514	0.8745869531	0.8845589757
6	accuracy	0.7279448748	0.7637396827	0.7593261674
	recall	0.7242937684	0.6813206226	0.682821326
	ROC-AUC	0.81111269	0.8076593131	0.803650409
	sensitivity	0.898022604	0.909957625	0.9072210416
	specificity	0.9008772016	0.8737037703	0.8673094064
7	accuracy	0.7303542164	0.7573645115	0.7608072907
	recall	0.6846199963	0.6645293662	0.6696588844
	ROC-AUC	0.7866151333	0.7921017494	0.7974703312
	sensitivity	0.8686082893	0.8910831894	0.8987133503
	specificity	0.8702003029	0.860175584	0.8615770936

Table G.2: Experiment 2 averaged results for 10 simulation runs of stations 8-16 for 6 epochs/round and 3 training rounds.

station	metrics	round 1	round 2	round 3
8	accuracy	0.7302799463	0.7592176596	0.7686835304
	recall	0.6688471794	0.6857015193	0.6610589847
	ROC-AUC	0.7758552909	0.8028371831	0.7949320897
	sensitivity	0.8586595178	0.8987488846	0.8924262747
	specificity	0.8519785166	0.8705853124	0.8616769984
9	accuracy	0.7287706137	0.7618014887	0.7598144487
	recall	0.7316322724	0.701682955	0.7113870904
	ROC-AUC	0.8060334524	0.8122168556	0.8147493899
	sensitivity	0.8957654834	0.9115092307	0.9155808985
	specificity	0.8867778381	0.8778053671	0.8805415183
10	accuracy	0.7314224044	0.7684105294	0.7661349944
	recall	0.7122083505	0.6859012587	0.6880258322
	ROC-AUC	0.8002511395	0.8116816112	0.8096043042
	sensitivity	0.8897481693	0.9156114459	0.9149315868
	specificity	0.8801637292	0.8761543887	0.873326557
11	accuracy	0.7337158248	0.771158857	0.7636989951
	recall	0.695121944	0.6754873225	0.7044605613
	ROC-AUC	0.7941710576	0.8058642319	0.8179849982
	sensitivity	0.8873599246	0.9103493861	0.9235332906
	specificity	0.8685419708	0.8691948567	0.8809468647
12	accuracy	0.7338454723	0.7565850094	0.7601981078
	recall	0.7009063482	0.7116314098	0.7038411754
	ROC-AUC	0.7961192012	0.8142135665	0.8109873704
	sensitivity	0.885196352	0.912386708	0.9117824861
	specificity	0.8737771273	0.8818725422	0.8775584613
13	accuracy	0.7301744297	0.7678371891	0.7594654229
	recall	0.6661127284	0.6494219601	0.6720616553
	ROC-AUC	0.7766267508	0.7895726636	0.7957060933
	sensitivity	0.8584537655	0.884031795	0.8901734087
	specificity	0.8585466743	0.8600533456	0.8662120567
14	accuracy	0.7288438678	0.7589759297	0.7617309615
	recall	0.6788483709	0.6777713961	0.6853819713
	ROC-AUC	0.7814488709	0.7997958991	0.806437254
	sensitivity	0.8611430228	0.8991639482	0.9088167697
	specificity	0.8650394306	0.8656944103	0.8693955317
15	accuracy	0.7350347102	0.758596162	0.76620958
	recall	0.6996739626	0.6918180519	0.6806008369
	ROC-AUC	0.7936375737	0.805972139	0.8048596904
	sensitivity	0.8891942263	0.9095896151	0.9097578004
	specificity	0.8647495329	0.8686992129	0.867823258
16	accuracy	0.7338614	0.7467972338	0.7666684687
	recall	0.6987206274	0.7209011217	0.682311217
	ROC-AUC	0.7946205338	0.8150785168	0.8073242505
	sensitivity	0.8827701343	0.9156236947	0.9116604129
	specificity	0.8719882634	0.8804704746	0.8709576428

Table G.3: Experiment 2 averaged results for 10 simulation runs of stations 17-24 for 6 epochs/round and 3 training rounds.

station	metrics	round 1	round 2	round 3
17	accuracy	0.7283976227	0.7435438275	0.764784847
	recall	0.7076677307	0.7211501598	0.6838886312
	ROC-AUC	0.7993539497	0.8120397568	0.807003711
	sensitivity	0.8898562342	0.9119488835	0.912733904
	specificity	0.8751517758	0.8802803636	0.8707591976
18	accuracy	0.7293542283	0.7631184816	0.7651801194
	recall	0.6794812083	0.6939255476	0.6889054775
	ROC-AUC	0.782495984	0.8099408388	0.8079333901
	sensitivity	0.8764579637	0.9144350171	0.9101427708
	specificity	0.8569584319	0.8722616315	0.8721143944
19	accuracy	0.735018803	0.765894711	0.7634030183
	recall	0.6956500014	0.7010546497	0.7032562693
	ROC-AUC	0.7940578991	0.8159252405	0.8146331906
	sensitivity	0.894655731	0.921380622	0.9201093713
	specificity	0.8668829004	0.8794662271	0.876892474
20	accuracy	0.7269577725	0.7539568841	0.7569432457
	recall	0.7122351357	0.7147129377	0.7048777209
	ROC-AUC	0.7988006813	0.8164231181	0.8126553827
	sensitivity	0.8855092355	0.9174641172	0.9158027636
	specificity	0.8794394135	0.8822716971	0.8779936433
21	accuracy	0.7294230163	0.7621243	0.7673267424
	recall	0.6941831633	0.6829207912	0.6720297039
	ROC-AUC	0.788244307	0.8036134392	0.8007042706
	sensitivity	0.8831683099	0.9079826772	0.9053630431
	specificity	0.8572983742	0.8666282967	0.8652688464
22	accuracy	0.7307400227	0.766682165	0.762263453
	recall	0.7192639351	0.7008954201	0.7081656218
	ROC-AUC	0.8021932244	0.8155358945	0.8170255125
	sensitivity	0.8935020089	0.9190010684	0.9202990234
	specificity	0.8775195956	0.8820375204	0.8809944451
23	accuracy	0.7340665119	0.7578043714	0.7672001521
	recall	0.6953368698	0.6913209483	0.6698507965
	ROC-AUC	0.7904948848	0.8066573292	0.8013328811
	sensitivity	0.8781191622	0.9072734714	0.9042940338
	specificity	0.868536268	0.8752961382	0.8692044516
24	accuracy	0.7272664383	0.7584168911	0.7622490898
	recall	0.6810897514	0.6855311334	0.6852869391
	ROC-AUC	0.7839112058	0.8038638175	0.8055695295
	sensitivity	0.875457868	0.9026251435	0.9053723961
	specificity	0.8600174561	0.8725392699	0.8725088909

Appendix H

Data Usage

Muñoz Sabater, J., (2019, 2021) [34] was downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store. Our results contain modified Copernicus Climate Change Service information from 2020. We have reduced the number of features contained in the information, the geographic area represented, and the number of years available. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains. Acknowledged as per [11].

The Fire Incident Locations - Historical data set was downloaded from the British Columbia Data Catalogue and is used under the Open Government License [40].

The 2021 Census Boundary file was downloaded from Statistics Canada and is used under the Statistics Canada Open Licence. Source: Statistics Canada, 2021 Census Boundary file, 27.09.2022. Reproduced and distributed on an "as is" basis with the permission of Statistics Canada.

BC Wildfire Service weather station data was downloaded from the Pacific Climate Impacts Consortium [8] and is modified and presented here under their terms of use [9].