Wildfire Air Quality Prediction:
A Data-Driven Approach

Subhankar Dhar, San Jose State University, USA*
https://orcid.org/0000-0003-1493-451X

Jerry Gao, San Jose State University, USA

ABSTRACT

Wildfires are extremely harmful to the environment. While producing gaseous pollutants and particles that cause smoke, wildfires also release carbon dioxide (CO_2), a greenhouse gas that will continue to warm the planet after the wildfire ends. This article delves into the impact of wildfires and air quality on human living conditions. The authors' machine learning models use wildfire data to forecast air quality with detailed indexes and geographic information during a wildfire. The work evaluates the performance of each machine learning model via statistical metrics like mean absolute error (MAE), R-squared (R2), and root mean squared error (RMSE). The experimental results used neural networks to predict a specific value for carbon monoxide (CO), ozone, and PM2.5. These are both promising and accurate, providing meaningful insight into air quality within a region. This work will be useful for cities, governments, citizens, and public safety.

KEYWORDS

Air Pollution, Air Quality Prediction, Big Data Analytics, Disaster Response Management, Environment Management, Machine Learning, Neural Network, Wildfires

INTRODUCTION

Wildfires can increase air pollution and cause severe damage to air quality by emitting carbon dioxide (CO2), carbon monoxide (CO), and other greenhouse gases that contribute to global warming and environmental hazards (Castelli et al., 2020). In addition, they can damage forests that otherwise remove CO2 from the air and inject aerosols into the atmosphere. Smoke from wildfires can cause serious health disorders and respiratory diseases (Camia & Amatulli, 2009; de Groot et al., 2007) by reducing the size of green forests and eliminating CO2 in the air (Reid et al., 2019). Pollutants like particulate matter (PM) 2.5, ozone (O3), nitrogen dioxide (NO2), and CO are major parameters in the air quality index (AQI) of a region (Tian et al., 2011).

DOI: 10.4018/IJDREM.330148 *Corresponding Author

This work delves into the impact of wildfires on air quality, using data science tools and technologies to predict the impact of air quality after a wildfire event. To achieve this goal, the authors must define key measurements with respect to air. For example, air has several important measurement metrics that affect air quality, including the CO level, ozone level, and PM2.5 (small inhalable particles with diameters no more than 2.5 micrometers) index. These pollutants can have severe negative impacts on human health, including wheezing, coughing, sore eyes, and throats. The authors, therefore, develop machine learning models to understand the health effects caused by wildfires and strengthen public awareness surrounding such events.

First, the authors develop a data-driven machine learning framework to predict air quality during and after wildfires. In addition, the study addresses the limitations of existing methods. The research provides insight into the significant impact of wildfires, enabling people and governments to prioritize environmental protections and prevention efforts. Residents can also monitor air quality changes more closely and prepare for potential wildfires.

Second, the authors use two sets of machine learning models to predict air quality during wildfire seasons. The study provides detailed explanations on how to determine and tune parameters. Analysis helps the public understand and anticipate air quality changes. It also enables individuals to compare and improve upon existing models.

Third, the study includes results driven by statistical analysis and machine learning. It aims to prove the significance of wildfires and predict air quality with improved accuracy. Descriptive and inferential statistical analysis tools, along with visualization diagrams and graphs, depict the difference between air quality before and after wildfires. The study's regression and deep learning models generate precise metrics to increase accuracy.

Fourth, the study presents a data-driven approach for deploying an online wildfire air quality alert system. It uses real-life data to develop a proof-of-concept implementation. This prototype demonstrates the novel functionalities of the authors' framework, benefiting both the public and city officials.

The article is organized as follows. The next section discusses related works. Then, the article describes the geographical location of the study, comprehensive data parameters, data pre-processing, and training and test data preparation. Wildfire air quality prediction models and experimental results are explored before navigating the system design, implementation framework, and visualization tools. Lastly, the study discusses the conclusion and future work.

RELATED WORK

Preisler et al. (2015) highlighted the shortcomings of research when predicting wildfire impacts from PM2.5 concentration at ground-level monitors in California. While most researchers rely on satellite-based observational tools, this work combined models with an autoregressive statistical model, incorporating weather and seasonal factors to identify thresholds for predicting unusual events. The study focused on ground-based monitoring of PM2.5 levels, with data consisting of hourly values of PM2.5 and meteorological data. Data was gathered from the United States Department of Agriculture Forest Services. Unexpectedly, the study found that smoke plumes could identify seasonal wildfire influence with high accuracy.

Reid et al. (2015) and Jaffe et al. (2013) evaluated the contribution of O3 caused by wildfires in the atmosphere in western U.S. The studies developed several statistical models to estimate the maximum daily eight-hour average (MDA8) O3 emissions and created methodology functions for three western areas in the U.S. The residual of the statistical model can provide information on O3 emissions that cannot be explained by normal wildfires.

Similarly, Preisler and Westerling (2007) developed a statistical model for forecasting fire-danger and producing one-month ahead wildfire-danger probability in the western U.S. The model predicted a monthly average temperature and drought severity index to demonstrate significant potential

Related Content

Crowdsourcing Investigations: Crowd Participation in Identifying the Bomb and Bomber from the Boston Marathon Bombing
Andrea H. Tapiaand Nicolas J. LaLone (2014). *International Journal of Information Systems for Crisis Response and Management (pp. 60-75).*
www.irma-international.org/article/crowdsourcing-investigations/129606

Open Infrastructure for a Nationwide Emergency Services Network
Mark Gaynor, Sarah Friedeck, Alan Pearce, Scott Bradnerand Ken Post (2011). *Crisis Response and Management and Emerging Information Systems: Critical Applications (pp. 96-111).*
www.irma-international.org/chapter/open-infrastructure-nationwide-emergency-services/53990

STAR-TRANS Modeling Language: Risk Modeling in the STAR-TRANS Risk Assessment Framework
Dimitris Zisiadis, George Thanos, Spyros Kopsidasand George Leventakis (2013). *International Journal of Information Systems for Crisis Response and Management (pp. 45-59).*
www.irma-international.org/article/star-trans-modeling-language/81274

Sociotechnical Uses of Social Web Tools during Disasters
Liza Potts (2014). *Crisis Management: Concepts, Methodologies, Tools, and Applications (pp. 531-541).*
www.irma-international.org/chapter/sociotechnical-uses-of-social-web-tools-during-disasters/90734

Preparing for Refugee Exodus in Crisis: Poland Case Study
Magdalena Denhamand Scott Vautrain (2018). *Handbook of Research on Environmental Policies for Emergency Management and Public Safety (pp. 166-188).*
www.irma-international.org/chapter/preparing-for-refugee-exodus-in-crisis/195194