

Idea #4: Predicting Wildfire Smoke Health Impact on Communities

Acceptance Likelihood: 4.3/5 — Strong Pick

The Problem

Canada's wildfire seasons are intensifying. The 2023 season was the worst on record, burning 18.5 million hectares. But the most widespread health damage comes not from the fires themselves — it comes from **smoke**. Wildfire smoke contains fine particulate matter (PM2.5) that penetrates deep into lungs and enters the bloodstream.

Key Statistics

- **2023 wildfires:** 49–400 premature deaths from acute PM2.5 exposure, 660–5,400 from chronic exposure
- **Respiratory ER visits** spike 5–15% during major smoke events
- PM2.5 from 2023 fires was detected **across the entire continent** — even NYC had hazardous air
- Canada has **260 air monitoring stations** (NAPS) collecting continuous data since 1969
- Indoor PM2.5 exposure more strongly associated with acute respiratory outcomes than outdoor
- Children, elderly, and those with pre-existing respiratory conditions are most vulnerable

Why It Matters

Unlike wildfires (which affect specific regions), **smoke affects entire populations** — millions of Canadians breathe hazardous air for weeks each summer. Predicting PM2.5 levels and their health impact enables proactive public health responses: air quality warnings, mask distribution, vulnerable population outreach, and hospital surge preparation.

Why Machine Learning Can Help

PM2.5 concentrations during wildfire events depend on complex interactions:

- **Fire location, size, and intensity**
- **Wind patterns and atmospheric conditions**
- **Temperature and humidity**
- **Distance from fire to population center**
- **Terrain and geographic features**
- **Historical air quality baselines**

ML excels at learning these nonlinear relationships to predict exposure levels.

ML Tasks

1. **Regression:** Predict daily PM2.5 concentrations from weather + fire data
2. **Classification:** Categorize air quality health risk levels (low/moderate/high/very high)
3. **Correlation analysis:** Link PM2.5 predictions to health outcomes (ER visits, hospital admissions)

Published Approaches

- **Random Forest:** Estimated hourly PM2.5 at 5km resolution in BC during fire seasons
- **Ensemble ML (RF + XGBoost):** Achieved high accuracy for structured pollution datasets

- **LSTM networks:** Captured temporal dependencies in pollution time series
 - **CNN:** Detected spatial patterns in pollution data
-

Datasets

1. National Air Pollution Surveillance (NAPS) Program (Primary)

- **Source:** Environment and Climate Change Canada via Open Canada
- **URL:** <https://open.canada.ca/data/en/dataset/1b36a356-defd-4813-acea-47bc3abd859b>
- **Contains:** Continuous monitoring of CO, NO2, O3, SO2, PM2.5, PM10 from 260 stations in 150 communities
- **History:** Data since 1969
- **Format:** Downloadable, well-structured

2. Air Quality Health Index (AQHI) Forecasts

- **Source:** Environment and Climate Change Canada via Open Canada
- **URL:** <https://open.canada.ca/data/en/dataset/a563e47d-6eb9-4f7f-933c-222ae49fe57f>
- **Contains:** AQHI scores (1-10 scale) based on O3, PM2.5, NO2 risk combination
- **Format:** JSON API and CSV — ideal for ML pipelines
- **Use:** Target variable for health risk classification

3. Canadian Wildfire Information System

- **Source:** Natural Resources Canada
 - **Contains:** Fire location, perimeter, intensity data
 - **Use:** Feature engineering — distance and intensity of active fires relative to monitoring stations
-

Proposed ML Pipeline

1. Data collection & integration
 - Download NAPS PM2.5 historical data
 - Fetch AQHI records via API
 - Obtain weather data (temperature, wind, humidity) from Environment Canada
 - Get fire location/intensity data from CWFIS
2. Feature engineering
 - Temporal: hour of day, day of year, season, rolling PM2.5 averages
 - Weather: wind speed/direction, temperature, humidity, precipitation
 - Fire: nearest active fire distance, total fire area within radius, fire intensity
 - Geographic: station latitude/longitude, elevation, urban/rural
3. Model training
 - Random Forest regressor (primary – predict next-day PM2.5)
 - XGBoost regressor (comparison)
 - Logistic regression classifier (AQHI risk category)
 - Train on non-fire season as baseline, evaluate on fire season
4. Evaluation
 - MAE/RMSE for PM2.5 prediction
 - Classification accuracy for AQHI categories
 - Compare fire season vs. non-fire season performance

- Analyze prediction lead time (how far ahead can we predict?)

5. Deliverable

- PM2.5 forecast dashboard for Canadian cities
- Health risk alert system prototype
- Analysis of which communities face highest cumulative smoke exposure

Strengths

- **Different from past wildfire projects:** Past cohorts predicted fire occurrence — this predicts health impact of smoke. Completely different ML task and dataset.
- **Excellent government data:** NAPS (260 stations since 1969) + AQHI with JSON API. No access barriers.
- **Timely:** Post-2023 wildfire crisis. Every Canadian experienced it.
- **Broad population impact:** Smoke affects millions, not just fire-adjacent communities
- **Clear actionable output:** Early warnings, hospital preparation, mask distribution
- **Strong ML fit:** Time-series prediction is classic ML task

Risks & Mitigations

Risk	Mitigation
Reviewers conflate with past wildfire projects	Explicitly differentiate in proposal: "health impact of smoke, not fire detection"
Health outcome data (ER visits) may be hard to access	Focus on PM2.5 prediction as primary. Health correlation as secondary analysis.
Complex data integration	Start with single-city model (Vancouver or Toronto), expand if time permits
Seasonal data limitation	2023 and 2024 provide extreme smoke events for model training

References

- ACS ES&T Air (2024): "Enhancing Wildfire Smoke Exposure Assessment: A ML Approach to Predict Indoor PM2.5 in BC" — <https://pubs.acs.org/doi/10.1021/acsestair.4c00204>
- Nature (2025): "Long-range PM2.5 pollution and health impacts from the 2023 Canadian wildfires" — <https://www.nature.com/articles/s41586-025-09482-1>
- PMC (2024): "Health Impact Analysis of Wildfire Smoke-PM2.5 in Canada (2019-2023)" — <https://pmc.ncbi.nlm.nih.gov/articles/PMC12852501/>
- Open Canada: NAPS dataset — <https://open.canada.ca/data/en/dataset/1b36a356-defd-4813-acea-47bc3abd859b>
- Open Canada: AQHI forecasts — <https://open.canada.ca/data/en/dataset/a563e47d-6eb9-4f7f-933c-222ae49fe57f>