World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Review Article)

Smog Prediction in Punjab Using Machine Learning Approaches: A Literature Review

Ibrahim Faisal Deen *

International Research Institute of North Carolina, North Carolina, USA.

World Journal of Advanced Research and Reviews, 2025, 28(03), 1331-1335

Publication history: Received 27 October 2025; revised on 16 December 2025; accepted on 18 December 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.28.3.4178

Abstract

Smog in Punjab is a severe public health and environmental concern, driven by a combination of crop residue burning, urban emissions, and adverse meteorological conditions. Fine particulate matter and toxic gases accumulate during winter months due to low temperatures, high humidity, and temperature inversions, leading to repeated episodes of hazardous air quality. This literature review synthesizes recent studies on the environmental drivers of smog and its impacts on public health and the economy. The review highlights the role of post-harvest crop burning as the primary contributor to particulate emissions and examines how meteorological factors exacerbate pollution levels. Evidence from existing research demonstrates a clear association between smog events and increased respiratory and cardiovascular hospitalizations, particularly among vulnerable populations. Emerging studies indicate that machine learning models can offer predictive capabilities by integrating real-time environmental and anthropogenic data, enabling early warning systems and targeted interventions. The findings underscore the urgent need for region-specific forecasting systems that can mitigate health risks, guide policy decisions, and inform proactive public health strategies in Punjab.

Keywords: Smog; Punjab; Machine Learning; Public Health; Crop Burning

1. Introduction

Smog is a hazardous blend of fine particulate matter ($PM_{2.5}$ and PM_{10}) and toxic gases like ozone, carbon monoxide, sulfur dioxide, and nitrogen oxides formed when pollutants from vehicles, industries, and fossil fuel burning interact with sunlight [1]. Smog has become a recurring public health emergency in Punjab, particularly during winter months. The province's air quality regularly exceeds World Health Organization safety limits, with $PM_{2.5}$ concentrations reaching 10-15 times the recommended threshold (Pakistan Air Quality Initiative), driven by vehicular emissions, industrial activity, and agricultural residue burning that contribute to photochemical reactions producing ground-level ozone [1].

A critical factor exacerbating Punjab's smog is post-harvest crop burning, where farmers set fire to rice stubble to clear fields for wheat planting, a practice shown to release massive amounts of particulate matter [2]. Satellite data reveals over 30,000 fire events annually in Punjab, peaking in October and November. When combined with Punjab's winter conditions of high humidity, and temperature inversions [3], these fires create smog episodes linked to increased hospitalizations for cardiovascular and respiratory diseases [4].

Current smog alerts in Punjab rely on retrospective air quality indices or simplistic statistical models, which fail to account for real-time crop fire data, cross-border pollution transport, or the nonlinear interactions between weather and emissions. Machine learning offers a promising alternative, as demonstrated by Siddiqui et al. [5] who successfully predicted PM10 levels using meteorological data and traffic patterns to generate actionable alerts. However, no study has yet developed a machine learning model tailored to Punjab's unique geographic and agricultural context.

* Corresponding author: Ibrahim Faisal Deen

This study investigates whether a machine learning model can be effectively developed to forecast smog levels in Punjab by integrating diverse environmental and anthropogenic factors, including humidity, temperature, and crop fire incidents. By examining recent literature, this paper aims to synthesize insights across two core themes that collectively inform the feasibility and potential design of such a model.

2. Methods

To establish a comprehensive understanding of smog dynamics in Punjab, I conducted a systematic literature review using Google Scholar as the primary database. The search employed targeted keyword combinations to capture diverse aspects of smog formation and impacts. Initial broad searches using terms like "smog causes AND smog effects" provided foundational knowledge about pollutant chemistry and health consequences, while geographically focused queries such as "Punjab AND smog AND public health" narrowed results to regionally relevant studies. To investigate specific environmental drivers, I used terms like "smog AND crop fires" and "smog AND humidity AND Lahore", which yielded papers on agricultural burning and meteorological influences. Finally, "machine learning AND smog prediction" helped identify technological solutions for forecasting.

The review adhered to strict inclusion and exclusion criteria to ensure relevance. Included studies focused on Punjab or comparable regions (e.g., Delhi with similar crop-burning practices) and were published within the last decade to reflect current conditions. Excluded were studies older than 10 years, as outdated data might not account for recent climate or policy changes. This filtering ensured the analysis remained aligned with Punjab's unique smog profile, where crop fires, winter meteorology, and cross-border pollution interact.

Through this process, two dominant themes emerged that directly address my research question: (1) public health and economic impacts, and (2) temperature, humidity, and crop fires as smog drivers. The health-related studies consistently reported spikes in respiratory/cardiovascular hospitalizations during smog season, linking them to $PM_{2.5}$ and ozone exposure. These findings underscored the urgency of predictive systems to mitigate health crises. Meanwhile, environmental studies revealed how Punjab's low winter temperatures and high humidity trap pollutants, while crop fires provide the primary emissions.

3. Results

Table 1 Summary of key peer-reviewed studies on smog, air pollution, and machine-learning-based prediction in Pakistan

Title (Citation)	Keywords	Themes / Focus Area	✓ Peer Reviewed	Database Used
[1] Arif	Smog, air pollution, health effects, Lahore	Causes and effects of smog in Punjab; basic overview of pollutants	✓	Google Scholar
[2] Rashid, Diyan, et al.	Crop burning, PM2.5, Punjab, soil health, health impacts	Link between stubble burning and smog formation in Punjab	✓	Google Scholar
[3] Mirza et al.	GIS, smog mapping, meteorology, Lahore	Monitoring smog patterns using GIS and meteorological data	✓	Google Scholar
[4] Iram et al.	Public health, respiratory illness, air pollution, Pakistan	Systematic review on health effects of smog and air pollution	✓	PubMed
[5] Siddiqui et al.	Machine learning, PM10 prediction, IoT, early warning	ML-based smog forecasting using meteorological data	✓	Google Scholar

4. Discussion

4.1. Public Health and Economic Impact

The severe smog episodes in Punjab present a growing public health crisis, with far-reaching consequences for the population's well-being. The toxic mix of pollutants, including high concentrations of particulate matter and hazardous gases, has been directly linked to a sharp rise in respiratory and cardiovascular emergencies across the region [4].

Medical facilities in urban centers like Lahore report alarming spikes in hospital admissions during peak smog seasons, particularly for conditions such as asthma exacerbations, chronic bronchitis, and acute cardiac events [4]. Vulnerable groups, including children, the elderly, and those with pre-existing conditions, face disproportionate risks, with long-term exposure contributing to reduced lung function and increased susceptibility to life-threatening illnesses [4].

Beyond immediate health effects, the smog crisis in Punjab carries significant socioeconomic burdens. Increased hospitalizations strain healthcare infrastructure, while productivity losses occur due to illness-related absenteeism and restricted outdoor activities [4]. The pervasive nature of air pollution means that even short-term exposure can have cumulative health consequences, emphasizing the urgent need for proactive mitigation strategies [4]. Emerging technologies, such as predictive modeling linked to early warning systems, could help reduce these impacts by enabling timely public health advisories and targeted interventions [5]. By integrating real-time pollution data with healthcare preparedness plans, authorities can better protect vulnerable populations and alleviate pressure on medical services during high-risk periods [4].

The development of smog prediction systems holds particular promise for safeguarding public health. Advanced forecasting could enable preemptive measures, such as redirecting traffic away from high-pollution zones or issuing health alerts to at-risk communities [5]. Such systems would not only reduce acute health emergencies but also support long-term policy efforts to curb pollution sources [4, 5]. However, their success depends on continuous monitoring, public awareness campaigns, and collaboration between environmental and health agencies to ensure data translates into actionable health protections [4]. Without decisive action, Punjab's smog crisis will continue to undermine population health and place unsustainable burdens on its healthcare system [4].

4.2. Temperature, Humidity and Crop Fires effects

The interplay between environmental conditions and anthropogenic activities creates a perfect storm for smog development in Punjab. While photochemical reactions between vehicular and industrial emissions generate ozone-rich smog during summer months [1], Punjab's winter smog follows a distinct pattern driven by three critical factors: temperature inversions, elevated humidity, and widespread crop residue burning [2].

During the cooler months, the combination of stagnant air and temperature inversions acts as a lid over the region, trapping pollutants near the surface [1]. This effect is exacerbated when humidity levels rise, as water molecules interact with airborne particles to form a dense, persistent haze [1]. The problem reaches its peak when these meteorological conditions coincide with the post-monsoon crop burning season [2]. Agricultural fires release massive quantities of particulate matter and precursor gases, which become suspended in the stagnant, moisture-laden air mass [2].

Low temperature and high humidity conditions create an environment where pollutants accumulate rather than disperse, with crop fire emissions serving as the primary ignition source for prolonged smog episodes [2]. The resulting pollution cocktail, a mix of fine particulates from burning biomass and secondary pollutants formed through atmospheric reactions, poses significant challenges for both air quality forecasting and mitigation efforts [1, 2].

The seasonal nature of these events suggests that predictive models must account for the nonlinear relationships between meteorological variables and emission sources [2]. For instance, while high humidity typically correlates with worse air quality, its impact is magnified when combined with active fire counts [2]. This complex interplay underscores the need for machine learning approaches that can capture these multidimensional interactions and provide accurate smog forecasts for public health protection.

4.3. Further discussion

The severe smog episodes in Punjab are now well-documented as a complex interplay of environmental and anthropogenic factors. Research has firmly established that crop residue burning during the October-November period serves as the primary emission source, releasing particulate matter and precursor gases that interact with winter meteorological conditions [2]. Multiple studies confirm that low temperatures, high humidity, and stagnant winds create an atmospheric environment where these pollutants accumulate rather than disperse, leading to prolonged smog events [3]. The public health consequences are equally well-characterized, with hospital data showing clear spikes in respiratory and cardiovascular emergencies during peak smog seasons, disproportionately affecting vulnerable populations [4]. This established understanding of smog formation and its health impacts provides a solid foundation for predictive modeling efforts.

However, significant gaps remain in the current literature. While numerous studies have examined either the environmental drivers of smog or its health effects in isolation, few have systematically analyzed how real-time monitoring of these factors could enable proactive interventions. Furthermore, there is limited research exploring how predictive systems could bridge the gap between environmental monitoring and public health response, particularly in resource-constrained settings like Punjab.

The feasibility of a machine learning approach to address these gaps appears promising based on existing evidence. The demonstrated success of similar models in urban environments [5] suggests that with appropriate modifications, these techniques could be adapted for Punjab's unique agricultural context.

The proposed machine learning model would use real-time data from multiple sources, including satellite fire detections, weather reports, and air quality readings, to forecast smog levels in Punjab. The system would "learn" patterns from past data, such as how crop fires and certain weather conditions (like low wind and high humidity) led to spikes in pollution.

Once trained, the model would take in fresh daily data and predict $PM_{2.5}$ levels for the next 1 to 3 days. If dangerous pollution levels are expected, it could automatically trigger health alerts via SMS or email to local authorities and residents.

While this report only outlines the concept, early research shows that such models have been successful in other cities. With proper data access and technical support, a similar system could be built and adapted to Punjab's unique environment.

Despite this potential, several limitations must be acknowledged. Current data collection systems in Punjab have notable gaps, particularly in rural areas where monitoring stations are sparse. This could lead to underestimation of true pollution levels during model training. Additionally, while machine learning can identify predictive patterns, it cannot account for sudden policy changes or agricultural practices that might alter emission patterns. The model's effectiveness would also depend on the quality and latency of input data.

Future research should focus on integrating real-time health data with environmental variables to improve model responsiveness. Expanding low-cost sensor networks in rural areas can address data gaps, while studying the effects of policy changes on emissions could enhance prediction accuracy. Finally, efforts should explore how to turn forecasts into actionable health alerts for local authorities.

5. Conclusion

The recurring smog crises in Punjab present a critical intersection of environmental degradation, public health risk, and technological opportunity. As the literature shows, smog in the region is driven by a combination of seasonal crop burning, meteorological conditions, and urban pollution sources. All of this disproportionately impacts vulnerable populations through elevated rates of respiratory and cardiovascular illnesses. While existing responses have largely been reactive, there is growing evidence that machine learning models could provide a more proactive and predictive approach.

By integrating diverse real-time datasets, including crop fire activity, weather conditions, and air quality metrics, such models have the potential to forecast smog events before they occur. This would enable early warning systems and targeted interventions to protect public health. However, realizing this potential requires addressing significant data and infrastructure gaps, especially in rural monitoring and policy integration.

Ultimately, this feasibility analysis suggests that a Punjab-specific smog prediction model is both possible and urgently needed. With interdisciplinary collaboration and investment in data systems, machine learning can move from concept to implementation, offering a powerful tool to combat one of the province's most pressing environmental challenges.

Compliance with ethical standards

Acknowledgments

The author would like to thank Samantha Bjelis for research mentorship and Dr. Daniel Jaffe for oversight and valuable feedback throughout the development of this literature review. Appreciation is also extended to the International Research Institute of North Carolina for academic support.

References

[1] Arif F. Smog: causes, effects and preventions. Annals of King Edward Medical University. 2016;22(4). doi:10.21649/akemu.v22i4.1456

[2] Rashid D, Ather MA, Fareed A, et al. A review of stubble burning-induced smog: ripple effects on soil ecosystems and human health. Agricultural Sciences Journal. 2025;7(1):44–65. doi:10.56520/asj.v7i1.451

[3] Mirza AI, et al. Environmental monitoring of smog in Lahore metropolitan area using GIS technology. International Journal of Innovative Science and Technology. 2022;4(3):809–820.

[4] Iram S, Qaisar I, Shabbir R, et al. Impact of air pollution and smog on human health in Pakistan: a systematic review. Environments. 2025;12(2). doi:10.3390/environments12020046

[5] Siddiqui SA, Fatima N, Ahmad A. Smart air pollution monitoring system with smog prediction model using machine learning. International Journal of Advanced Computer Science and Applications. 2021;12(8):401–409. doi:10.14569/IJACSA.2021.0120846