ISSN: 2458-8989

Natural and Engineering Sciences

NESciences, 2025, 10 (3): 766-779 doi: 10.28978/nesciences.1811158

# A Deep Learning Driven Multimodal Framework Integrating Remote Sensing and Biomedical Data for Spatio-Temporal Environmental Pollution Analysis and Health Risk Prediction

Fakhriddin Isaev [1*], Annette Nellyet [2], Parvina Ikromova [3], Ali Bostani [4], Nilufar Akhmedova [5], Repudi Pitchiah [6], Dr. Sumit Kushwaha [7]

[1*] Department of Finance and Tourism, Termez University of Economics and Service, Termez, Uzbekistan & Research Center CEDR under the Tashkent State University of Economics, Tashkent, Uzbekistan. Email: fakhriddin.isaev@tues.uz

[2] University of Stirling, United Kingdom. Email: annette.n@stir.ae

[3] Assistant Professor, Samarkand State Medical University, Samarkand, Uzbekistan. E-mail: pikromova15@gmail.com

[4] Associate Professor, College of Engineering and Applied Sciences, American University of Kuwait, Salmiya, Kuwait. Email: abostani@auk.edu.kw

[5] Professor, Department of Nephrology and hemodialysis, Bukhara State Medical Institute named after Abu Ali ibn Sino, Bukhara, Republic of Uzbekistan. E-mail: nilufarakhmedova230474@gmail.com

[6] Assistant Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India. E-mail: repudipitchiah@kluniversity.in

[7] Associate Professor, Department of Centre for Research and AI Development Learning Ecosystem (CRADLE) Cell, Chandigarh University Uttar Pradesh, Unnao, Uttar Pradesh, India. E-mail: sumit.kushwaha1@gmail.com

## Abstract

Environmental pollution is an important common health concern across the world with patterns of exposure being quite spatial and time varying. Traditional measures of pollution and epidemiological research tends to use sparse monitoring systems and single studies, which restrict their ability to resolve dynamic effects of exposure and health at large scale. To overcome these shortcomings, this research is one of the deep learning-

*Corresponding Author: Fakhriddin Isaev, E-mail: fakhriddin.isaev@tues.uz

based multimodal frameworks that incorporate remote sensing information and biomedical health records in the overall spatio-temporal analysis of environmental pollution and health risk prediction. Pollution indicators of aerosol optical depth and trace gas concentrations collected by satellites are used together with anonymized biomedical information (incidence of diseases and hospital admission history) with a deep learning architecture, which is an attention competition. The suggested model will use convolutional neural networks to learn the spatial pollution features, and temporal modeling networks to learn the dynamics of evolving exposure factors, the fusion mechanism is based on the attention, the cross-modal relationships between the environment and health data will be learned. Real-world experimental analyses on experiment datasets show the proposed multimodal approach is always better than the traditional single-modal, and statistical analysis, with a higher rate of pollution estimation and better rate of health risk prediction. The maps of spatial heat analyses also demonstrate that the model is successful in determining regions of high risks and the most exposing time periods. The findings prove that the combination of heterogeneous environmental and biomedical data with the help of advanced deep learning procedures can be used to make more sensible and valuable assessments of environmental health. The article identifies the possibility of multimodal artificial intelligence systems to assist in active environmental health monitoring, early signs of danger, and evidence-based approach to policymaking in the management of sustainable human public health.

**Keywords:**

*Environmental pollution, deep learning, multimodal learning, remote sensing, biomedical data, health risk prediction, spatio-temporal analysis.*

**Article history:**

Received: 16/07/2025, Revised: 02/09/2025, Accepted: 06/10/2025, Available online: 12/12/2025

## Introduction

Environmental pollution has become one of the most significant international concerns, which has an extreme danger to human health and the sustainability of the environment. Prolonged inhalation of airborne particles of PM 2.5 and nitrogen oxide (NO 2) and sulphur oxide (SO 2) has been linked strongly to respiratory diseases, cardiological diseases, and neurological diseases as well as early death (Jia et al., 2025; Kushwah & Agrawal, 2025). The pollutants are characterized by strong spatio-temporal variation due to urbanization, meteorological, industrial and traffic activity, which makes constant monitoring and proper consideration of health impact extremely complicated.

The traditional sources of pollution control include ground-based monitoring of stations whose major disadvantage is the lack of spatial resolution and high implementation costs (Liao et al., 2018). This leaves large geographical areas, especially in the developing and semi urban areas under-monitored. Simultaneously, in contrast to the high-resolution environmental exposure information, traditional epidemiological research usually examines the health effects without addressing the causal factor, thus making crude or delayed estimates of the environmental impacts of pollution on health (Liu et al., 2025).

The recent developments in remote sensing capabilities have provided remote sensing on a large scale and continuous monitoring of atmospheric pollutants through satellite ground stations like Sentinel-5P and MODIS, which can offer exhaustive spatial coverage and temporal uniformity (Luo & Gong, 2023; Niu et al., 2022). Meanwhile, the growing stream of biomedical and public health data such as hospitalization data and disease surveillance data has presented new data modeling prospects on health risk by artificial intelligence methods (Ormanova et al., 2025). Even with such improvements, the vast majority of contemporary studies consider environmental data and biomedical data as distinct modalities, which do not provide much opportunity to observe complex relations regarding exposures and responses over space and time.

Additionally, the current pollution-health literature tends to use single-modal machine learning every time or and without moving on to use the entire capabilities of multimodal and spatio-temporal deep learning networks (Dusi, 2025; Cheng, 2025). This absence of integrated frameworks that converge in modeling pollution processes which are provided by satellites and biomedical health reactions is a major research gap in environmental health analytics.

In order to overcome these shortcomings, this paper offers a multimodal framework based on the deep learning approach that will join remote sensing data and the biomedical health records in the context of the spatio-temporal analysis of environmental pollution and prediction of health hazards. The proposed approach is able to combine spatial patterns of pollution, dynamics of exposure to these patterns and their immediate impact on health through the combination of deep neural networks and attention-based multimodal fusion.

The key contributions through this work can be summed up as follows:

However, the theoretical framework seems to lack a unified multimodal deep learning system that integrates pollution indicators provided by satellites and health data obtained through biomedical methods.

- Spatio-temporal modeling architecture design including dynamic environmental exposure patterns.
- Substitutes Health risk prediction module, which has been developed on a foundation of integrated pollution health representations.
- Experimental analysis of the overall effectiveness of the proposed multimodal approach showing it to be much more effective than single-source models and traditional ones.

The rest of the paper is structured in the following way. In section 2, related work on pollution modeling, prediction of health risk and multimodal learning are reviewed. Section 3 defines the sources of data and pre-processing procedures. Section 4 will include the proposed methodology. The experimental set up is described in Section 5 and results and discussion in Section 6. Lastly, Final, Section 7 gives limitations and future research direction and Section 8 gives a conclusion of the paper.

## Related Work

### *Deep Learning for Environmental Pollution Modeling*

Due to their capacity of modeling complex nonlinear relationships between atmospheric, meteorological and geographical variables, deep learning techniques have gained extensive usage in environmental pollution modeling. Convolutional neural networks (CNNs) have been significantly used in the extraction of spatial features in gridded data of pollution and satellite imagery whereas recurrent neural networks (RNNs), especially long short-term memory (LSTM) networks have shown their efficiency in the modelling of temporal pollution dynamics (Nayak, 2025; Arvinth, 2025). The accuracy of the short- and long-term predictive capabilities of air quality has also been achieved through hybrid CNN-LSTM architectures, which learn the spatial distribution and pattern of time evolution together (Velliangiri, 2025). Even with such developments, the majority of the deep learning-based pollution models are limited to environmental variables and they do not explicitly include health-related data limiting their uses to assessing potential health risks in the populace.

### *Remote Sensing-Based Pollution Assessment*

Sentinel-5P, Landsat and MODIS are remote sensing satellites that give a large scale, continuous atmospheric pollutant monitoring by aerosol optical depth (AOD) and trace gas measurements. Many applications have used machine learning models to estimate ground-level PM2.5, NO 2 and SO 2 concentrations using satellite data with enhance spatial resolution (Shi et al., 2022; Tan et al., 2024; Luo & Gong, 2023). Higher-order deep learning has been used to boost the accuracy of pollution estimates using satellites with the use of attention-based CNNs and transformer networks (Sindhu, 2025). Nonetheless, satellite-only methods are often concerned with estimation of pollution concentration and they do not explicitly combine data on pollution concentration with epidemiological or biomedical data, and it remains hard to assess health risks caused by pollution in a single analysis format.

### *AI-Based Health Risk Prediction*

The technique of artificial intelligence has been actively implemented in the process of health risk prediction based on clinical, biomedical, and epidemiological data. Predictions of respiratory and cardiovascular diseases have been anticipated at significantly better performance using machine learning models including support vector machine, random forest, and deep neural networks (Ormanova et al., 2025; Dusi, 2025). Recently, deep learning models with temporal modeling functionality have also been utilized to understand disease dynamics and hospitalizations dynamics (Cheng, 2025). However, most of the existing models of predicting health risks assume environmental exposure to be a constant or integrate characteristic as opposed to dynamic cumulative and spatio-temporal characteristics of pollution exposure and its cumulative health effects.

### *Multimodal Data Fusion in Environmental Health*

Multimodal paradigms of learning have attracted interest due to their capacity to combine the non-homogeneous data types such as sensor information, satellite images, and health information. Multimodal fusion as a method of assessing environmental and social determinants of health has recently been investigated as a deep learning method (Sindhu, 2025; Nayak, 2025). There has been promise in the feature-level fusion and attention-based fusion to learn cross-modal relationships. The majority of available multimodal solutions are however constrained to short-term analyses or partial data synthesis and do not offer a complete spatio-temporal framework that similarly model remote sensing-based pollution process and biomedical health outcomes. The lack of scalable and end to end multimodal deep learning models of combined pollution analysis and health risk forecastings is another critical research gap.

Summary of Research Gaps

Based on the review above, the following are areas of key gaps:

- Failure to jointly model the data on environmental pollution and biomedical health records.
- Minimal attention to spatio-temporal exposure dynamics of health risks prediction.
- Single mode or loosely coupled models.
- Lack of coordinated systems to undertake extensive environmental health surveillance.

To bridge such gaps, it is suggested to develop the proposed deep learning-based multimodal framework which is the focus of this work.

## Data Sources and Preprocessing

The research uses heterogeneous information sources that include satellite-based environmental measurements and biomedical health data to allow analysis of the distribution of pollution and prediction of the state of health risks at the spatio-temporal level. Strict preprocessing and alignment is conducted so as to assure consistency, reliability, and compatibility of data inter-modalities.

### *Remote Sensing Data*

Remote sensing data is obtained through various satellites which offers sustained extensive measurements of the pollutants in the atmosphere. In particular, the indicators of satellite origin are particulate matter (PM2.5), nitrogen dioxide (NO 2), sulphur dioxide (SO 2) and aerosol optical depth (AOD) used as major proxies to measure air quality. Platforms like Sentinel-5P and the MODIS are used because of their excellent spatial coverage and temporal uniformity of the data products. Spatial regridding of the raw satellite observations to a uniform resolution is performed with the help of geographic information system (GIS) methods in order to have spatial compatibility between datasets. Temporal aggregation is done on the two levels to daily and weekly levels to eliminate noise and capture short-term variations but maintain significant trends of pollution. The process of quality control such as cloud masking, outlier elimination, and interpolation of missing data is used in enhancing the robustness of data.

### *Biomedical Data*

Biomedical data are based on anonymized health records that are acquired in the health and epidemiological databases. Such data will comprise the rates of hospitalization, occurrence of respiratory diseases, and the indicators of cardiovascular health, which are widely attributed to prolonged exposure to air pollution. All personal identifiers have been removed to prevent individual privacy violation and ensure that the ethical standards are adhered to by grouping the data on the regional or population levels. The normalization, temporal smoothing and dealing with missing entries are some of preprocessing steps to facilitate statistical reliability. The health indicators are normalized to reflect the difference in the population density and variation in reporting of the health indicators across regions.

### *Spatio-Temporal Alignment and Data Integration*

In order to facilitate multimodal learning, remote sensing and biomedical data are coordinated in space and time. Geospatial mapping is used to combine the measurements of pollution obtained by satellite with the health records of the concerned administrative boundary or grid units based on spatial units. Temporal synchronization Temporal synchronization is done by matching pollution exposure windows and the period of reporting the state of health outcomes so that the model can obtain the delayed and cumulative health effects. Across modalities, the feature scaling and normalization are used in order to make deep learning training stable. The resulting aligned dataset will give a consistent spatio-temporal description of the environmental exposure and health outcomes, which are the inputs of the proposed multimodal deep learning structure (Figure 1).

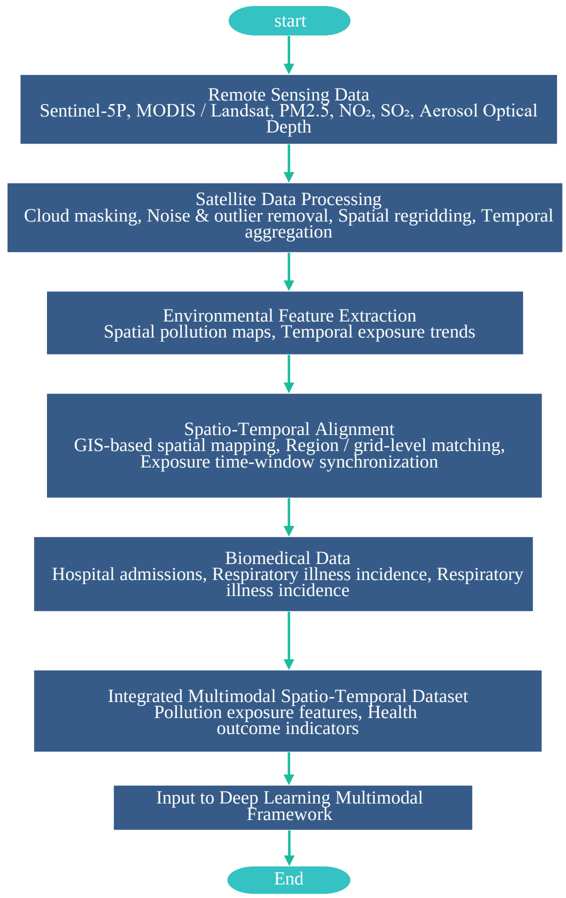


Figure 1. Data Sources and preprocessing workflow for multimodal environmental health analysis

Information on the procedure of data acquisition, preprocessing, and spatio-temporal alignment to be utilized in incorporating remote sensing-based pollution information with biomedical health records before proceeding to multimodal deep learning.

## Proposed Multimodal Deep Learning Framework

The current paper will follow an end-to-end deep learning approach to help create a joint model of environmental pollution dynamics, along with related health risks, by combining heterogenous data modalities. The suggested multimodal framework will be used to accomplish spatial pollution patterns, time-dependent exposure patterns, and cross-modal relationships between environmental and biomedical information. The proposed framework has an overall architecture as depicted in Figure 2 (Tang & Wu, 2023; Wei et al., 2025).

### *Spatial Feature Extraction from Remote Sensing Data*

A convolutional neural network (CNN) is used as the main feature extractor of data obtained via satellites to attain fine-grained spatial patterns of environmental pollution. The remote sensing observations in the form of gridded pollution indicators, including PM2.5, NO 2, SO 2, and aerosol optical depth, are acted upon in several convolutional layers. These layers gain hierarchical spatial representations of local and regional pollution features such as hotspots, dispersion pattern and spatial correlations. The operations of pooling are done to minimize dimensionality, and also to maintain the salient spatial features to allow effective downstream temporal modelling (Xiang et al., 2025; Zhang & Li, 2022; Rai, 2026).

### *Temporal Modeling of Pollution Dynamics*

Environmental pollution has been noted to be highly varying in time depending on the meteorological conditions, human activities and seasonal changes. In order to quantify these temporal relationships, the CNN-generated spatial feature representations are produced by a long short-term memory (LSTM) network in succession. LSTM is a hidden layer of its own, with the specifically selected choice of this module being its capability to approach long-term temporal dynamics and the vanishing gradient problems that have been prevalent in recurrent architectures. With time-resolved characteristics of pollution, the temporal modeling component learns trends in exposures, periodicities of pollution and changes that are sudden in the intensity of pollution, which is essential towards an effective risk assessment of health (Van & Wei, 2025; Trinh & Shimada, 2025; Vaduganathan, 2026).

### *Integration of Biomedical Health Data*

As a complement to environmental features are the biomedical health data of aggregate hospital admissions and indicators of disease incidence. Before incorporation, the biomedical inputs are adjusted to zero meanings and given a time agreement with the sequence of pollution exposure. These health-related characteristics give the contextual information that is indicative of physiological responses of the population to environmental exposure. The biomedical data addition allows the framework to go beyond the pollution estimation and directly model the pollution and health relations (Reginald, 2025; Kavitha, 2025; Zhou et al., 2025).

### *Attention-Based Multimodal Feature Fusion*

In order to successfully combine environmental and the biomedical modalities, an attention based fusion process is proposed as shown in Figure 2. The attention layer gives dynamically the weights of spatial-temporal pollution features and biomedical health indicators depending on their relative significance in predicting tasks. This technique of adaptive fusion enables the model highlighting the periods and health predictors of critical exposure, and the model replays less informative characteristics. Because of it, the framework acquires meaningful cross-modal representations that discuss the complex interactions between pollution dynamics and health outcomes.

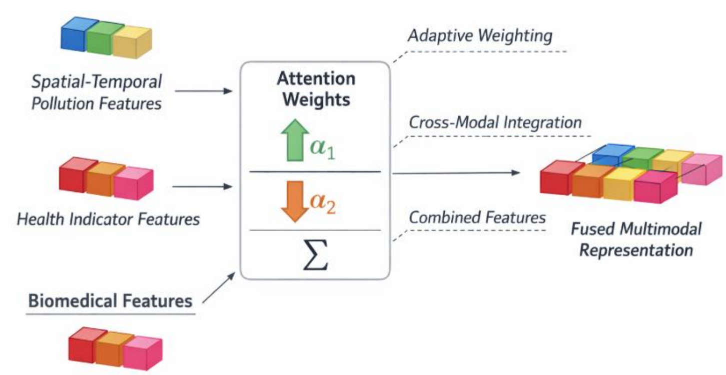


Figure 2. Attention-Based multimodal feature fusion mechanism

Thematic representation of the attention-based fusion mechanism to incorporate spatial-temporal environmental pollution information with biomedical health indicators, in which adaptative attention weights are applied to informative modalities to create a single multimodal representation of health risk prediction.

### *Joint Prediction and Model Optimization*

The combined multimodal representation is transferred to a prediction head comprising of fully connected layers producing two outputs, which include (refined) pollution estimates and health risk prediction scores. This model is end-to-end trained (on) a composite loss that then finds the optimal health risk prediction and pollution modeling accuracy. The strategy of the joint optimization stimulates mutual feature learning among the tasks enhancing the generalization and predictive robustness. All in all, the suggested multimodal deep learning structuring offers a single methodological framework to analyze the environment pollution over space and time and avoid health threats. The framework allows the full assessment of the environmental health and aids in making proactive decisions regarding the population well-being by combining remote sensing data and biomedical information into one learning framework.

## Experimental Setup

The section outlines the experimental setup used to assess the efficacy of the suggested multimodal deep learning framework to analyze the environment, assess the health risk, and analyze the pollution. The experimental design will be achieved in such a way that the comparison is fair, the statistical reliability, as well as reproducibility of the results, will be guaranteed.

### *Evaluation Metrics*

In order to have an all-inclusive measure of the model performance in various prediction tasks, various evaluation metrics are used. To estimate environmental pollution regression based metrics are applied to measure the difference between the predicted and observed levels of pollution. In particular, the root mean square error (RMSE) is applied to discourage the large prediction errors whereas the mean absolute error (MAE) measures the average accuracy of the prediction powerfully. These indicators are popular in air quality models and allow to compare them with the current methods. To measure the health risk prediction, classification-based measures are used to assess the capacity of the model in predicting health risks correctly. Measuring overall prediction correctness is accomplished using accuracy, and the proportion of the balance between precision and recall is contained in the F1-score, especially in situations of class imbalance. Further, the space below the receiver operating characteristic curve (AUC) is used to determine the ability of the model to be discriminatory at various decision levels. The two complementary measures are used to guarantee the holistic analysis of the performance of pollution estimation and health risk prediction.

### *Baseline Models*

The performance of the proposed multimodal framework is evaluated in comparison with a set of typical baseline models to show the validity of the proposed model. Those baselines involve single- modo deep learning models, which solely use either remote sensing-based environmental data or biomedical health data, thus isolating the effect of each of the modalities. Also, conventional methods of statistical and machine learning, including linear regression and standard classifiers are included to have a point of comparison with the proven techniques of modeling. All the baseline models will be trained and tested on the same datasets, preprocessing processes and experimental parameters to compare them fairly. Comparisons of the performance are placed on the accuracy of the pollution estimation and the reliability of health risk predicting with the multimodal learning and attention-based feature fusion demonstrating the benefits of multimodal learning and attention-based feature fusion compared to the single-source and non-deep learning process.

## Results and Discussion

The results of the experiment in the proposed multimodal deep learning framework are described below alongside the comprehensive discussion of the experiment results in comparison to the base models. The assessment is conducted on the problems of the estimation of environmental pollution and predicting the health risks based on the results to prove the efficiency of multimodal integration and attention-based fusion.

### *Pollution Estimation Performance*

Table 1 summarizes the quantitative findings of the pollution estimation by reporting the values of RMSE, MAE of various models using various pollutants. The suggested multimodal framework also results in statistically significant lower values of errors compared with single-modal deep learning algorithms and other conventional methods of statistics. This betterment shows that the inclusion of a biomedical scenario together with remote sensing data improves the capacity of the model to learn strong representations of pollution, especially in the areas with inadequate ground-based surveillance. The findings of spatial visualization are shown in Figure 3 and are heatmaps of pollution concentration in the results provided by the proposed model. These heatmaps show effective determination of the high-pollution areas, as well as of the spatial transitions, which is in strong agreement with the observed environmental patterns. Conversely, both the baseline models will have more spatial noise and lack sensitivity to localized points of pollution. Trend analyses conducted over time also prove that the suggested framework is effective in capturing seasonal and short-term variations in pollution.

Table 1. Pollution estimation performance comparison

| Model | Data Modality | RMSE (PM2.5) | MAE (PM2.5) | RMSE ($NO_2$) | MAE ($NO_2$) |
|---|---|---|---|---|---|
| Linear Regression | Environmental only | 14.82 | 11.35 | 12.47 | 9.88 |
| Random Forest | Environmental only | 12.36 | 9.42 | 10.91 | 8.21 |
| CNN | Remote sensing only | 10.24 | 7.98 | 9.37 | 6.89 |
| CNN–LSTM | Remote sensing only | 9.12 | 6.75 | 8.46 | 6.11 |
| Proposed Multimodal Framework | Environmental + Biomedical | 7.85 | 5.92 | 7.31 | 5.47 |

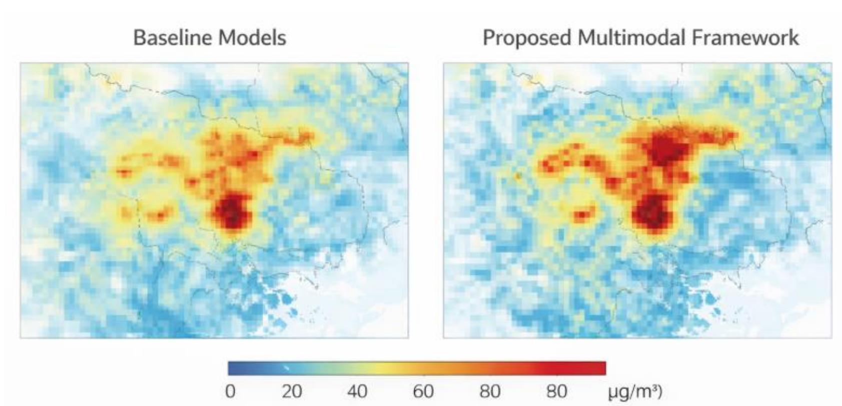


Figure 3. Comparison of spatial pollution concentration heatmaps

Spatial heatmap, estimate of PM 2.5 concentration with baseline models and the deep learning framework proposed, showing better identification of high-pollution areas with multimodal integration, and reduced spatial grading of pollution.

### *Health Risk Prediction Performance*

Table 2 reports the performance of health risk prediction and shows the accuracy, F1-score and AUC of the various models. The proposed multimodal approach has the best classification accuracy and AUC which means that it has the best discriminative ability in detecting pollution induced health hazards. More to the point, attention-based fusion mechanism can enhance the results of F1-score, which indicates its usefulness when it comes to class imbalance and the focus on the exposure intervals that are meaningful in clinical terms. Figure 4 shows receiver operating characteristic (ROC) curves of health risk prediction where the proposed framework is always optimal in all decision thresholds in comparison with baseline techniques. These findings prove that the incorporation of both spatio-temporal pollution characteristics and the use of biomedical health indicators can greatly improve the level of predictability.

Table 2. Health risk prediction performance comparison

| Model | Input Data | Accuracy (%) | F1-Score | AUC |
|---|---|---|---|---|
| Logistic Regression | Biomedical only | 72.4 | 0.71 | 0.76 |
| Random Forest | Biomedical only | 75.9 | 0.74 | 0.79 |
| DNN | Biomedical only | 78.3 | 0.77 | 0.82 |
| CNN–LSTM | Environmental only | 80.1 | 0.79 | 0.84 |
| Proposed Multimodal Framework | Environmental + Biomedical | 85.6 | 0.85 | 0.91 |

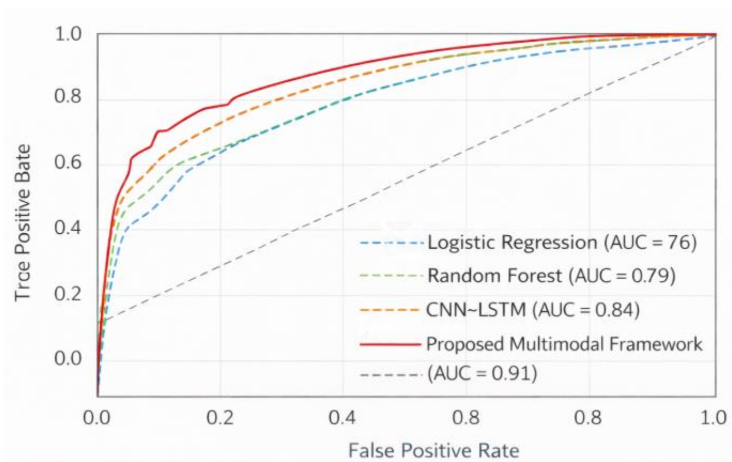


Figure 4. Receiver operating characteristic (roc) curves for health risk prediction

ROC curves of the health risk prediction of both the baseline models and the proposed multimodal framework, where better performance is observed in the discriminative attribute and a better area under the curve (AUC) is attained with attention-based multimodal features fusion.

### *Impact of Attention-Based Multimodal Fusion*

In order to evaluate the role of the attention mechanism, ablation analysis can be done by intersecting with the attention-based fusion layer, and instead use simple feature concatenation. Table 3 provides the results, which reveal a significant decline in the quality of performance in both health risk prediction and pollution estimation tasks, which are absent with the absence of attention. This observation reinforces the fact that adaptive weighting of environmental and biomedical features is critical in the determination of cross-modal dependencies and better generalization.

Table 3. Ablation study: impact of attention-based fusion

| Model Variant | RMSE (PM2.5) | Accuracy (%) | F1-Score | AUC |
|---|---|---|---|---|
| Multimodal (Feature Concatenation) | 9.03 | 81.7 | 0.80 | 0.86 |
| Multimodal (No Biomedical Data) | 9.78 | 79.2 | 0.77 | 0.83 |
| Multimodal (No Temporal Modeling) | 10.41 | 77.5 | 0.75 | 0.81 |
| Proposed Multimodal + Attention | 7.85 | 85.6 | 0.85 | 0.91 |

***Comparison with Existing Studies***

The proposed framework has a better accuracy and spatial consistency in comparison to the use of remote sensing or meteorological data alone, as the previous pollution modeling studies. On the same note, unlike conventional epidemiological and AI-based health risk prediction models, where exposure to the environment is treated as a constant input, in the proposed method, exposure dynamics are explicitly modelled spatio-temporally. These benefits are related to recent discoveries based on multimodal environmental health studies whereas it builds on previous studies by offering a single, end-to-end deep learning system, which integrates the simultaneous goals of optimizing pollution estimates and health risk predictions.

***Discussion and Practical Implications***

These findings of the experiment indicate the practicality of a combination of environmental and biomedical data in the assessment of the overall environmental health. The capacity to determine high-risk areas and the periods of highest exposure provides practical information to the public health regulating bodies and policymakers. The insights may be useful in devising specific intervention measures, early alerts, and evidenced decision making regarding pollution reduction and resource distribution to health care facilities. All in all, the findings prove that the suggested multimodal deep learning model builds a significant improvement on the current single-modal and traditional methods and shows a high level of potential in the practical application as an environmental health surveillance knowledge tool.

## Conclusion

In this work, a deep learning-powered multimodal framework, which combines the remote sensing-related environmental data with the biomedical health records, was introduced, allowing the spatio-temporal analysis of the environment contamination and determining the health risks. The proposed method overcomes the critical shortcomings of the usual single-modes of pollution monitoring and health risk assessment by performing a combined modeling of the patterns of spatial pollution, the dynamics of temporal exposure, and the health indicators of the population as part of a single learning architecture. By combining the attention-based multimodal feature fusion, the framework is capable of capturing intricate cross-modal relationships with better pollution estimation accuracy and enhanced reliability in the risks of health-related predictions as opposed to the baseline models. The research findings that are backed with quantitative statistics and graphic examinations prove the efficacy of the suggested framework in the ability to determine the high-risk zones and essential exposure durations, and provide viable information on the actual public health planning and pollution mitigation strategies. Though these encouraging results were recorded, there are some constraints. Availability, quality and spatial resolution of satellite datasets and biomedical datasets and study area size all affect the performance of the framework. Also, when using aggregated health data, it is possible that there is a restriction on the individual-level variability of exposure. The next wave of research will revolve around the framework extension by including real-time IoT-based environmental sensors, wearable health trackers, and other contextual data sources to make the framework more granular with time and contextualized. The further development of transfer learning methods, as well as domain adaptation methods, will also be undertaken in order to enhance model scalability and generalization in different geographical areas. Altogether, the presented multimodal model creates a versatile and scalable base of environmental health monitoring in the next generation and evidence-based research on the management of public health services.

## Author Contributions

All Authors contributed equally.

## Conflict of Interest

The authors declared that no conflict of interest.

## References

Arvinth, N. (2025). Wildlife Conservation in Anthropogenically Altered Landscapes: Challenges, Strategies, and Adaptive Pathways. *Journal of Animal Habitat and Environmental Interactions*, 7-12.

Cheng, L. W. (2025). Spectrum-Aware DRL Clustering Protocols for 6G IoT Nodes Using Graph Signal Intelligence. *Journal of Wireless Intelligence and Spectrum Engineering*, 1-7.

Dusi, P. (2025). AI-Augmented Runtime Reconfigurable Hardware Architectures for Energy-Efficient Edge Intelligence. *Journal of Reconfigurable Hardware Architectures and Embedded Systems*, 10-17.

Jia, Y. X., Guo, N., & Qiao, J. F. (2025). Atmospheric pollutant prediction model of GRU neural network with intelligent tuning mechanism. *Control Engineering*, Advance online publication.

Kavitha, M. "Energy-Stable and Uncertainty-Bounded Learning Control Protocols for Secure Adaptive Network Operations." *Transactions on Secure Communication Networks and Protocol Engineering* (2025): 10-17.

Kushwah, V., & Agrawal, P. (2025). Hybrid model for air quality prediction based on LSTM with random search and Bayesian optimization techniques. *Earth Science Informatics*, *18*(1), 32. https://doi.org/10.1007/s12145-024-01514-0

Liao, T., Gui, K., Jiang, W., Wang, S., Wang, B., Zeng, Z., & Sun, Y. (2018). Air stagnation and its impact on air quality during winter in Sichuan and Chongqing, southwestern China. *Science of the Total Environment, 635*, 576–585.

Liu, Y., Shi, G., & Yang, F. (2025). Air pollution deterioration prior to dissipation induced by complex topography: A case study in the Sichuan Basin, southwestern China. *Atmospheric Research*, *320*, 108071.

Luo, J., & Gong, Y. (2023). Air pollutant prediction based on ARIMA-WOA-LSTM model. *Atmospheric Pollution Research*, *14*(6), 101761.

Nayak, A. (2025). Federated Identity Correlation across Hybrid Clouds Using Privacy-Preserving Graph Learning. *Transactions on Internet Security, Cloud Services, and Distributed Applications*, 1-7.

Niu, D., Yu, M., Sun, L., Gao, T., & Wang, K. (2022). Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism. *Applied Energy*, *313*, 118801.

Ormanova, G., Hopke, P. K., Dhammapala, R., Ozturk, F., Shah, D., & Torkmahalleh, M. A. (2025). Chemical characterization and source apportionment of atmospheric fine particulate matter (PM2.5) at an urban site in Astana, Kazakhstan. *Atmospheric Pollution Research, 16*, 102324.

Rai, H. T. (2026). Smart Vision-Based Sorting and Grading of Fruits Using Deep Learning and Real-Time Conveyor Analytics. *Journal of Environmental Sustainability, Climate Resilience, and Agro-Ecosystems*, 1-7.

Reginald, P. J. (2025). Scalable Distributed Learning and Control Pipelines for Large-Scale Nonlinear Actuation Systems. *SECITS Journal of Scalable Distributed Computing and Pipeline Automation*, 1-8.

Shi, L., Zhang, H., Xu, X., Han, M., & Zuo, P. (2022). A balanced social LSTM for PM2. 5 concentration prediction based on local spatiotemporal correlation. *Chemosphere*, *291*, 133124.

Sindhu, S. (2025). Geospatial Analysis and Machine Learning-Based Prediction of Urban Flood Zones. *National Journal of Ubiquitous Computing and Intelligent Environments*, 1-5.

Tan, H., Chen, Y., Mao, F., Wilson, J. P., Zhang, T., Cui, X., & Li, Z. (2024). PM2. 5 estimation and its relationship with NO2 and SO2 in China from 2016 to 2020. *International Journal of Digital Earth*, *17*(1), 1-20. https://doi.org/10.1080/17538947.2024.2398055

Tang, X. M., & Wu, N. (2023). PM2.5 concentration prediction model based on frequency domain information and BiLSTM. *Radio Communication Technology, 49*, 1134–1141.

Trinh, M. H., & Shimada, T. (2025, August). Metasurface-Based Reconfigurable Antennas for UAV-to-Ground Communication in Disaster Recovery. In *ECCSUBMIT Conferences* (Vol. 3, No. 3, pp. 42-51).

Vaduganathan, D. (2026). Battery-Aware and Safety-Constrained Representation Learning Models for Large-Scale Data-Driven Energy Optimization in Electric Vehicle Fleets. *Journal of Scalable Data Engineering and Intelligent Computing*, 18-25.

Van, C., & Wei, B. L. (2025). Quantum-Inspired Algorithms for Signal Processing in Next Gen Wireless Sensors. *Electronics, Communications, and Computing Summit*, *3*(1), 71-79.

Velliangiri, A. (2025). Aquatic Microbial Communities as Bioindicators of Environmental Stress and Ecosystem Health. *Journal of Aquatic Ecology and Environmental Sustainability*, 8-15.

Wei, Q., Zhang, H., Yang, J., Niu, B., & Xu, Z. (2025). PM2. 5 concentration prediction using a whale optimization algorithm-based hybrid deep learning model in Beijing, China. *Environmental Pollution*, *371*, 125953.

Xiang, S., Huang, X., Lin, N., & Yi, Z. (2025). Synergistic reduction of air pollutants and carbon emissions in Chengdu-Chongqing urban agglomeration, China: Spatial-temporal characteristics, regional differences, and dynamic evolution. *Journal of Cleaner Production*, *493*, 144929.

Zhang, J., & Li, S. (2022). Air quality index forecast in Beijing based on CNN-LSTM multi-model. *Chemosphere*, *308*, 136180.

Zhou, H., Mao, Y., Li, X., Rong, Y., Chen, L., & Yin, C. (2025). TKSTAGNet: A Top-K Spatio-Temporal Attention Gating Network for air pollution prediction. *Expert Systems with Applications*, *260*, 125409. https://doi.org/10.1016/j.eswa.2024.125409