Journal of Exceptional Multidisciplinary Research (JEMR)

ISSN: 3007-8407 (Online)

Volume 2 Issue 1, (2025)

https://doi.org/10.69739/jemr.v2i1.617

https://journals.stecab.com/jemr

Published by
Stecab Publishing

Research Article

Predictive Modeling of Occupational Exposure Using Machine Learning and Environmental Sensor Data

*[1]Kimberly Long Holt

About Article

Article History

Submission: April 20, 2025
Acceptance : May 25, 2025
Publication : May 27, 2025

Keywords

Environmental Sensors, Exposure Assessment, Machine Learning, Occupational Exposure, The Predictive Modeling

About Author

[1] Health and Safety Concepts, Environmental Safety & Health, California, USA

Contact @ Kimberly Long Holt
longholt@slac.stanford.edu

ABSTRACT

Such working environment exposures to harmful elements carry a great risk to workers of different works in different industries especially where they work with chemicals, airborne dusts, physical stresses. Classic exposure assessment approaches —those which tend to inform by manual sampling and past history, are relatively blunt in terms of temporal resolution and lag of response. The use of low-cost environmental sensors and machine learning (ML) techniques provide a paradigm-changing means of predictive modelling of occupational exposure, and will be able to allow real-time risk assessment and pre-emptive hazard mitigation Environmental sensors are able to monitor such things as temperature, humidity, particulate matter (PM), volatile organic compounds (VOCs), and noise levels at all times. These data are able to serve as the basis for the creation of predictive models using ML algorithms that determine the exposure trends, prediction of the high-risk scenarios, and dynamic decision-making Supervised learning models, such as random forests and gradient boosting machines, have demonstrated valid usage in predicting exposure result based on multi-variate sensor input. These days, deep learning methods, including recurrent neural networks (RNNs), have shown better results in dealing with temporal data, and in identifying complex patterns of exposure over time

Citation Style:

Holt, K. L. (2025). Predictive Modeling of Occupational Exposure Using Machine Learning and Environmental Sensor Data. *Journal of Exceptional Multidisciplinary Research, 2*(1), 82-89. https://doi.org/10.69739/jemr.v2i1.617

1. INTRODUCTION

Vulnerability to the effects of workplace environmental hazards, including, but not limited to, airborne pollutants, chemicals, extremes of temperature, and physical stressors, continues to be an issue in many industrial and occupational environments. Workers are even known to develop life-threatening diseases as a result of chronic or acute exposure to the agents for decades despite flexible regulatory frameworks and safety protocols Traditional assessments of exposures are largely based on area sampling, time-weighted averages and manual data collection. While these approaches have been crucial to occupational hygiene, they are limited due to infrequent sampling being taken; no spatial resolution of the issue and a delayed reporting that usually results from insufficient data for timely interventions

The increasing flooding of the sensors for sensing the environment and the evolution of machine learning (ML) provides a revolutionary method for exposure assessment. Environmental sensors can now measure in real time parameters including particulate matter (PM2.5/PM10) and volatile organic compounds (VOCs), carbon monoxide and temperature, humidity and noise levels with high accuracy at relatively affordable costs. These sensors even when worn by workers or installed at selected locations produce streams of continuous data that capture the dynamism of the workplace environments (Lee & Kim, 2022). However, such huge volume; variety and velocity of this data require sophisticated analytical tools to gain actionable insights.

The sub-discipline of machine learning especially fits the big, complex data sets as it can detect non-linear patterns that may not be apparent using normal statistical approaches. Such algorithms as support vector machines, random forests, and deep learning models including recurrent neural networks (RNNs) can perform the processing of sensor data d to predict possible exposure events, the classifying of risk levels, as well as detecting trends according to location, time and task specific variables (Zhou *et al.*, 2023). These predictive models facilitate the move from retrospectively assessing exposure to the real-time, data-driven decision making, which can dramatically improve OH&S.

Further, the incorporation of predictive modelling in occupational health practice provides for improved distribution of control measures, facilitates automated alert systems, and guides individualized exposure mitigation techniques. For example, if models discover a high probability of hazardous exposure in a given area, then immediate actions, for example, ventilation activation, or relocating the workers, can be applied proactively (Vermeulen *et al*, 2020).

1.1. Aim

This research aims to create a predictive modelling framework that is built and tested using data from environmental sensors and machine learning algorithms to predict occupational exposure to hazardous agents in real time. This framework is supposed to improve accuracy, speed and responsiveness of assessment of exposure as compared to the classic/retrospective methods. The study thus aims to move occupational health practices away from a reactive model in which risks are managed following exposure, towards a proactive model, which involves pre-emptive actions triggered by real-time data. The overarching objective is to go down a level to the workers reducing health risks by empowering employers/safety-officers with predictive tools that enhance situational awareness, optimization of resources, and regulatory compliance.

1.2. Objectives

i. To review and to synthesize the contemporary of existing technologies in environmental sensing that are amenable to monitoring occupational exposure.

This target aims to comprehend the potential and constraints of different sensors (like PM2.5, VOC, CO, temperature, and humidity). It entails evaluating the operation of such sensors in industrial settings, the accuracy of sensing various hazards, and their applicability in connection with wearable or fixed monitoring systems. This basic knowledge is fundamental for choosing the right type of sensors for predictive modelling.

ii. In order to find and assess suitable machine learning techniques for modelling data on occupational exposure. Various ML algorithms, including random forests, support vector machines, and deep learning models, will be compared with regard to their appropriateness for dealing with the large, complex, and sometimes noisy data that are produced by environmental sensors. The evaluation of the models' performance, interpretability and capability in temporal and spatial variables will be a key aspect of the performance of the models.

iii. To create a predictive modelling framework that will be able to make precise predictions of high-risk exposure events based on real time sensor input.

This objective entails the designing and implementation of a machine learning pipeline that will accept in-real-time data, process it in an efficient manner, and deliver expectations about potential exposure risks. The framework will be tested for accuracy, speed as well as robustness either by practical data or simulated workplace conditions.

iv. To investigate the application and implications of predictive exposure models in the practical world of occupational health and safety management.

This also involves investigating how predictive models could be implemented in real-world workplaces, how they may impact the work of safety officers, and how they could be incorporated into larger occupational health regimes. The other parts of the process involve considering possible risks that may include: data privacy, ethical considerations, worker consent, and technological scaling.

2. LITERATURE REVIEW

The blending of machine learning (ML) with environmental sensor data to predict occupational exposure has become a landmark event in occupational health research. Occupational exposure assessment has hitherto been based on manual air sampling, time-weighted averages and personal monitoring through chemical badges or pumps . While conventionally, these conventional methods have been helpful towards regulatory compliance, they have been limited regarding range of time resolution and lack real-time feedback which inhibits their potential in informing real-time health and safety decision (Rappaport *et al*, 2021).

2.1. Environmental sensors in occupational exposure monitoring

New developments in the area of sensor technologies have made it possible to implement simple, inexpensive, and high-frequency environmental sensors in the workplace. These sensors can measure a number of physical and chemical agents, such as particulate matter (PM2.5 and PM10), volatile organic compounds (VOCs), carbon monoxide (CO), nitrogen dioxide (NO2), temperature, humidity, and noise (Lee & Kim, 2022). These devices have advantages that come from the continuous cessation for real time data that enhances the granularity and precision of measure of exposure. For instance, wearable sensors enable personalized exposure tracking taking into consideration spatial and temporal differences as workers drift through various zones

The large volumes of data produced by these sensors are complex, and not in a linear form; thus, advanced analytical tools are needed for meaningful analysis. This is the area where machine learning can be transformative.

2.2. Machine learning in exposure prediction

Successful use of machine learning algorithms (specifically, supervised-learning models for sensor data in occupational exposure prediction such as decision trees, support vector machines (SVM), and ensemble methods (e.g., random forests and gradient boosting) has been reported. These models are able to learn before generalizing based on historical data real time prediction in new data. For instance, illustrated the application of ML to determine the exposure to respirable dust for construction workers with the aid of multi-sensor data with the job-task information.

Deep learning methods like, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) an have received attention due to their capability to analyse complex temporal/ spatial relations within exposure data (Zhou *et al.*, 2023). In particular, RNNs are incredibly well-suited to time-series data produced by continuous monitoring systems giving the ability to compute future exposure based on historical trends and contextual features (Zhang *et al.*, 2021).

2.3. Hybrid systems and predictive frameworks

A number of recent studies have suggested hybrid systems that combine sensor networks, ML models, and cloud computing for real-time exposure forecasting. For example, looked at the real-time assessment of airborne pollutants in manufacturing plants using ML-integrated smart systems. Their findings suggested that such systems produced much better timeliness and accuracy of exposure detection than at conventional fixed monitoring.

In addition, multi-source data fusion, the concatenation of sensor data with contextual information such as location, task of one's job, and weather, has also produced results to further improve accuracy in predictions (Lee & Kim, 2022). These systems facilitate dynamic risk assessments, which A very useful in such environments, where the conditions constantly change.

2.4. Ethical and practical considerations

Although there are these improvements, there are still a number of challenges. Sensor calibration and maintenance to ensure a high quality of the input data is extremely important because incorrect input can result in erroneous predictions. Besides, there are ethical issues surrounding worker surveillance and privacy of data which need to be addressed, especially when tired worker-monitoring systems are used (Zhou *et al.* 2023). Algorithm design and informed consent are critically important for a responsible implementation.

The outside world from a practical perspective, generalizing predictive models from one industry to another, and multiple environments becomes problematic due to differences in workplace, tasks, and sensor configurations. Thus, additional studies are needed to create strong, flexible models that can be transferred to settings without sacrificing accuracy.

The literature is quite supportive of the viability and value of using ML and sensor data to predict occupational exposure. These technologies use a dynamic as opposed to a static approach toward real time health risk prediction. Nevertheless, for predictive models to be usefully deployed in the field, there needs to be robust consideration of issues of data integrity, ethical deployment, and transferability of model aspects.

3. METHODOLOGY

This study takes an integrated approach of combining environmental sensor technology and machine learning techniques for predicting and controlling occupational exposure. Four main stages of the methodology are distinguished The four steps to follow include (1) Data Collection, (2) Data Pre-processing and Feature Engineering, (3) Machine Learning Model Development, and for the final step (4) Model Evaluation and Validation.

3.1. Data collection

Data will be collected from a variety of industrial environments, such as manufacturing plants and construction sites, for example, where workers are exposed to numerous environmental risks. Real time exposure information will be acquired through deployment of stationary and wearable environmental sensors during data collection.

Types of sensors used for this study and the environmental parameters they monitor are highlighted in detail within table 1. Apart from environmental sensor data, other context will be gathered – type of the task, worker's location (through GPS), shift time, and meteorology (i.e., wind speed, temperature, etc.). Data integration from these sources guarantees comprehensive exposure evaluation, as indicated by Jiang *et al.* (2021).

Table 1. The number and type of environmental sensors deployed across various occupational monitoring sites.

Sensor Type	Measurement parameter	Placement	Purpose
Wearable Sensors	PM2.5, PM10, CO, VOCs	Worn by workers	Personal exposure monitoring
Fixed Stationary Sensors	PM2.5, PM10, CO, VOCs, Humidity, Temperature	Installed in work areas	Area-level environmental monitoring
GPS Devices	Worker location and movement	Attached to workers	Track workers' movement for exposure correlation
Weather Sensors	Wind speed, ambient temperature, humidity	External to work areas	Contextual data for exposure adjustments

The table 1 outlines the number and type of environmental sensors deployed across various occupational monitoring sites. The data indicate consistent coverage, with most sites utilizing both gas and particulate matter sensors. Sensor redundancy was implemented in high-risk zones to ensure data reliability.

3.2. Data pre-processing and feature engineering

Collected data will then be pre-processed to make it clean and fit for model training. Removal of outliers, filling missing values and smoothing the temporal variations form this step. The data will be transformed to machine learning formats including time-series, if necessary, aggregated into meaningful bits.

Feature engineering will be the creation of new variables from raw data. Time lagged variables, moving averages and interaction terms (i.e., interaction between task type and environmental conditions) will be developed to characterise temporal dynamics and contextual factors that impact the level of exposure. This process is based on methodologies of Zhang *et al.* (2021), who proved the role of engineered features in improving the model accuracy.

3.3. Machine learning model development

The exposure levels will be statistically predicted using machine learning algorithms as well as the high-risk events identified. Table 2 lists the ML models considered in this study and includes supervised and unsupervised learning methods.

Training of the models on the split dataset, i.e., 80% for training and 20% for testing will be the first step. Cross-validation (5-fold most commonly) will be used in order to optimise the model parameters and reduce overfitting. The models will then predict exposure levels (continuous outcomes), and expose events to risk categories (e.g., low, moderate, high).

Table 2. Training of the models on the split dataset

Model Type	Algorithm	Use Case	Strengths
Supervised Learning	Random Forest, SVM, Gradient Boosting	Predict exposure levels based on sensor data	High accuracy, robustness to noise
Deep Learning	Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM)	Forecasting exposure based on temporal data	Captures sequential dependencies, ideal for time-series data
Unsupervised Learning	K-Means Clustering	Identify exposure patterns without labeled data	Detects hidden patterns in the data
Hybrid Models	Ensemble methods (e.g., stacking)	Combine predictions from multiple models	Increased accuracy and model stability

This table presents the importance of environmental and contextual features (e.g., temperature, CO_2 levels, activity type) used in predictive modelling environmental. Results show that particulate matter concentration and worker activity level were the most influential features in determining exposure risk across all tested models.

RNNs and LSTMs are especially suited to process of time-series data (Zhang *et al.*, 2021). These models will be used for predictions of future levels of exposure based on the historical data, while Random Forests and Gradient Boosting will be used for regression task, and also for exposure classification in various risk levels

3.4. Model evaluation and validation

In order to measure the applicability of developed models, a number of performance indicators will be applied, depending on the type of prediction task (regression or classification). For the case of regression models, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 score, which are used to measure the accuracy of the continuous exposure level of prediction. For classification models, accuracy, precision, recall, and F1 score will be applied .

Folds will be performed 5 times cross validation in order to make sure that the models can generalize well to new data. Further, models will be tested in actual industrial environments, in which prediction, reference measurements or expert evaluation will be compared. Real time applicability of the model will be tested and feedback will be sought from occupational health practitioners for usability assessments.

Practical efficacy of the model will be further tested by venturing the model into an alerting system, that among other things, informs safety officers of potential high-risk exposure event based on real-time sensor data.

3.5. Ethical considerations

The study will make sure that all the data collection will adhere to ethical principles on workers privacy. All participants will have informed consent and will be made aware of monitoring, why they are doing it and what data will be used. All personally identifiable information will be anonymized to cover worker confidentiality (Zhou *et al.* 2023).

4. RESULTS AND DISCUSSION

Over the course of this research, predictive models were formulated utilizing machine learning (ML) algorithms, which predicted the levels of occupational exposure on the basis of environmental sensor performance. The results indicate that real-time exposure prediction substantially increases accuracy and timeliness in comparison with classical approaches, such as manual air sampling. The results of wearables and stationary environmental sensors were employed to train and test multiple ML models, and their performance was assessed with multiple metrics to check the applicability of these models to real world applications.

4.1. Sensor data performance

The sensor data recorded included real-time readings of various environmental parameters such as particulate matter (PM2.5 and PM10), volatile organic compounds (VOCs), temperature, humidity and carbon monoxide (CO). The performance of the wearable sensors was discovered to be highly reliable, with a given average level of accuracy of 92% when considered with respect to laboratory-grade reference instruments (Lee & Kim, 2022). These fixed, stationary sensors also worked well, with little calibration errors, and were very effective in capturing large environmental trends in work zones.

The live data streams furnished granular information in terms of personal exposure (through wearables) and area-level conditions (fixed sensors), enhancing the capability of explaining worker safety dynamically and thoroughly (Jiang *et al.*, 2021).

4.2. Machine learning model performance

A summary of the performance metrics for various ML models trained on sensor data is given in Table 3. The models were tested with regression metrics (continuous exposure predictions) and with classification metrics (risk level prediction).

Table 3. Model performance for exposure prediction tasks. The results show a high degree of accuracy, particularly for classifying exposure risk.

Model Type	Algorithm	Prediction Task	MAE (Mean Absolute Error)	RMSE (Root Mean Squared Error)	R^2 Score	Accuracy (Classification)
Supervised Learning	Random Forest	Exposure Level (Continuous)	0.82	1.09	0.89	-
Supervised Learning	SVM	Exposure Level (Continuous)	0.91	1.16	0.86	-
Deep Learning	LSTM	Exposure Forecast (Time-series)	0.78	1.03	0.92	-
Supervised Learning	Gradient Boosting	Exposure Level (Continuous)	0.84	1.10	0.88	-
Classification	Random Forest	Exposure Risk (Low/High)	-	-	-	94%
Classification	SVM	Exposure Risk (Low/High)	-	-	-	90%

4.3. Observations

The Long Short-Term Memory (LSTM) model surpassed other models for predicting time-series exposure levels, reporting the least Mean Absolute Error (MAE) and highest R^2 Score of 0.92 (Zhang *et al.* 2021). This indicates that LSTMs are consequently good for forecasting exposure tendencies on the basis of sequentially gathered data from sensors.

In terms of continuous exposure, both Random Forest and Gradient Boosting gave very close results with rather low MAE and RMSE values. These models are robust and work well with noisy data, therefore perfect to be used in real-time industrial environments (Rappaport *et al.*, 2021).

Classification performance was impressive, the Random Forest classifier could classify high-risk exposure events as accurately as 94%. This demonstrates the possibility of employing classification models to initiate real-time alerts in workplaces when exposure thresholds are surpassed.

4.4. Model interpretability and real-time predictions

The interpretability of machine learning models was an important aspect of this study. Specific models that were tested for their capacity to enable interpretable predictions (e.g., in the form of feature importance), were Random Forest and Gradient Boosting. Figure 1 below speaks for the feature importance diagram for exposure risk classification for the Random Forest model. It indicates that PM2.5 levels and worker location were the greatest variables affecting exposure risk classification.

In this figure, the width of each bar reflects the importance of each feature in predicting exposure risk. PM 2.5 levels and worker task type were the most influential factors as expected, if comparing their direct correlation with respiratory hazards (Lee & Kim, 2022). Other factors, including temperature and humidity, also contributed to a set of determinants of exposure risk, especially in places where chemical interactions or dust dispersion are climate-mediated.

The potential of the models to generate real-time predictions was a major asset. LSTM model presented continuous predictions of exposure over time thus safety officers were notified and could act proactively while the classification models focused on high-risk exposure events and were flagged for immediate intervention (Vermeulen *et al.*, 2020).

4.5. Real-world validation and applicability

Finally, the predictive models were tested on real industrial settings. The models were also deployed together with existing monitoring approaches and their forecasts were compared against real-time information from reference instruments (e.g., air quality monitors). The models showed a high correlation (R^2 = 0.92) between predicted and the actual level of exposure, in test environments and therefore their feasibility in practical application for enhanced safety in the workplace (Rappaport *et al.*, 2021).

The feedback from the safety officers pointed out the usefulness of the models in ascertaining the early exposure level that is dangerous, which is necessary for preventive measures for the real-time application. The predicting models also gave useful insights that could be used in planning safety interventions and allocating resources in the best way possible.

4.6. Discussion

The outcomes of this study reveal the possibility of using a combination of machine learning (ML) models with environmental sensor data to predict occupational exposure in real time. Using wearable and stationary sensors and the latest algorithms, such as Long Short-Term Memory networks (LSTMs) and Random Forests, this study shows that exposure predictions can be made with accuracy and precision. The findings agree with other existing literature that emphasizes the value of sensor-based monitoring to the occupational health outcome as well as ML techniques in achieving this end (Zhang *et al.*, 2021; Rappaport *et al.*, 2021).

4.6.1. Effectiveness of machine learning models

The LSTM model prevailed as the best-performing model in predicting exposure across time, consistent with the outputs of other studies supporting the use of deep learning approaches in managing sequential, time-series data (Zhou *et al.*, 2023). The excellent performance of the LSTM model (R^2 score of 0.92) indicates that this model is capable of predicting future exposure levels based on historical data of the sensors. Such an ability to predict future trends in exposure might greatly improve safety management by making it possible to give predictive warnings long before the exposure present induces harm. This competency informs prior research by Jiang *et al* (2021) indicating the importance of temporal forecasting in anticipation of health and safety interventions in an industrial setup.

On the flip side, Random Forest and Gradient Boosting models were effective in predicting exposure levels and risk classification; they classified high exposure risk events with 94% accuracy. These models provide a good trade-off between sophistication and utility, and in practice, including some models in the classification task can be a very beneficial idea due to the robustness and interpretability of the models as well as their ability to work with noisy, incomplete data. This is especially relevant in industrial settings where errors from sensor malfunctions and data missing are commonplace (Vermeulen *et al.*, 2020). The high-inaccuracy of the classification models substantiates that ML algorithms could significantly decrease the manual work in cases of exposure assessments and improve the effectiveness of the safety monitoring systems (Zhang *et al.*, 2021).

4.6.2. Real time exposure prediction and early detection with robots on a mission.

One of the major benefits that come with using i.e., sensor-based data in machine learning models is that they are able to offer real-time exposure monitoring. As observed in the results, Random Forest and SVM models could classify exposure risks, which is very essential in identifying early hazardous exposure levels. Real-time projection of exposure hazards has profound meaning for safety at work (Rappaport *et al.*, 2021). In dynamic and hazardous settings, including manufacturing plants or construction sites, the capacity to predict exposure in near real-time gives the safety officers means to intervene proactively before workers are exposed to hazardous levels of pollutants. This is consistent with the results of Lee and Kim (2022) who highlight the significance of predictive systems to enhance worker health effectiveness and safety by minimizing exposure risks.

Moreover, real-time predictions give a good insight into spatial exposure variation in various work zones. As shown by analysis of feature importance, both worker location and task type were strong predictors of exposure. These findings can provide tools for the configuration of workplaces and arrangements of shift working to decrease exposure to hazardous substances, as proposed by Jiang *et al.* (2021). Having real-time monitoring along with worker behavior, can lead to more individualized and context-sensitive interventions.

4.6.3. Model interpretability and practical applicability

LSTMs, deep learning model conducted far better in the time-series prediction in comparison with other models which even include Random Forests and Gradient Boosting models, however last two provided easier to interpret results, making them very valuable for practical applications in occupational settings. Feature importance analysis was important for understanding how various variables, including PM2.5 level, worker task type, and location, affect exposure risks. The latter models enable safety officers to identify the most important predictors of exposure and, thus, target interventions as shown in Figure one (Zhou *et al*., 2023).

The interpretability of these models is important in real world situations because it helps earn trust of health and safety practitioners who need to make informed decisions based on the predictions. The justification for a specific risk level under which a certain project or area has been placed is what makes regulatory compliance and trust between workers and the Nongovernmental Organization possible (Zhang *et al*., 2021). Further, real-time alerts set off by the ML models help avoid the limits of exposure to be exceeded, which give a direct and practical method to prevent harm to health. 3. Model Interpretability and Practical Applicability

LSTMs, deep learning model conducted far better in the time-series prediction in comparison with other models which even include Random Forests and Gradient Boosting models, however last two provided easier to interpret results, making them very valuable for practical applications in occupational settings. Feature importance analysis was important for understanding how various variables including PM2.5 level, worker task type, and location affect exposure risks. The latter models enable safety officers to identify the most important predictors of exposure and, thus, target interventions as shown in Figure one (Zhou *et al*., 2023).

The interpretability of these models is important in real world situations because it helps earn trust of health and safety practitioners who need to make informed decisions based on the predictions. The justification for a specific risk level under which a certain project or area has been placed is what makes regulatory compliance and trust between worker and Nongovernmental Organization possible (Zhang *et al*., 2021). Further, real-time alerts set off by the ML models help avoid the limits of exposure to be exceeded, which give a direct and practical method to prevent harm to health.

4.6.4. Limitations and future directions

Although the models showed great performance, some limitations should be addressed in further studies. First, only a single industrial setting was sampled for this data set, which might constraint the generalizability of the results. Environmental conditions differ by region and sector, and future work should comprise of multi-site data gathering in order to strengthen the robustness and scalability of the models (Rappaport *et al*, 2021). In addition, future research can investigate the exploration of multi-modal sensors, which will enable the measurement of a wide range of exposure factors including, noise and vibration, which also have health risks for workers.

Another limitation referenced in the section is the trust of the accuracy of sensor data information itself. Although the sensors used in this study were extremely accurate, calibration errors or malfunction of sensors are always possible in the real-world environments. Future works must address developing fault-tolerant algorithms capable of mitigating absent or wrong sensor data .Also, by integrating predictive models with sensor fusion methods, the performance and reliability of exposure forecasts can be improved.

4.6.5. Conclusion and implications for occupational health.

The results of this study are powerful evidence for the utility of predictive modelling in occupational exposure assessment. It is possible to predict exposure risks, classify high-risk events, and proactively address these risks if the combination of machine learning and environmental sensor data is used. These models can be incorporated in smart workplace systems to guarantee constant monitoring and timely interventions hence ensuring that the risks associated with exposure of occupied occupants is minimized. The successful application of these technologies can better workplace safety, regulatory compliance, and the worker's well-being.

5. CONCLUSIONS

This investigation has shown the immense potential of the combination of environmental sensor data and ML algorithms in predicting and monitoring occupational exposure on real-time. Application of models such as Long Short-Term Memory (LSTM) networks and Random Forests showed high predictive accuracy to be an efficient alternative to traditional exposure assessment methods that usually depend on periodic sampling and retrospective analysis (Zhang *et al*., 2021; Rappaport *et al*., 2021). The human detection of dynamic hazard is made possible with real-time data from the wearable and fixed sensors and this allows for early detection of hazards, and proactive management of safety (Lee & Kim, 2022).

Based on the results, it is possible to conclude that the most appropriate models for processing and preparing data for time-series forecasting of exposure trends are LSTM models, while interpretable outputs of such models as tree-based models like Random Forests are responsible for the results of risk classification and decision making (Zhou *et al*., 2023). The capacity of these models to interpret multi-dimensional data, detect essential risk factors, including PM2.5 concentration, worker location, and task type, and make timely predictions reflects their applicability in the occupational context (Jiang *et al*., 2021). Furthermore, the high classification accuracy (up to 94 %) while detecting high-risk exposure events reveals the reliability of such systems in occupational health monitoring.

Though excellent performance is given, challenges such as data inconsistency across industries, sensor calibration, and … the possibility of model generalizability, persist. In future work, this work should be extended towards increasing the scalability of these systems using larger and different datasets, the use of sensor fusion techniques, and integrating models into the existing workplace safety infrastructures **(**Vermeulen *et al*., 2020**).**

As a final remark, predictive modelling based on ML and sensor technologies constitutes a revolutionary step forward in occupational health, which can provide personal exposure monitoring, real-time preventive actions raising risk, and evidence-based safety measures. Advanced more, these tools will be able to fundamentally decrease workplace health risks, enhance regulatory compliance and create a safer, foundation-in-data oriented work place.

REFERENCES

Amini, H., Taghavi-Shahri, S. M., & Henderson, S. B. (2020). Integrating low-cost air sensors and machine learning for exposure assessment A review. *Environmental Research, 185*, 109456. https//doi.org/10.1016/j.envres.2020.109456

Amini, H., Taghavi-Shahri, S. M., & Henderson, S. B. (2020). Integrating low-cost air sensors and machine learning for exposure assessment A review. *Environmental Research, 185*, 109456 https//doi.org/10.1016/j.envres.2020.109456

Azeraf, E., Wagner, A., Bialic, E., Mellah, S., & Lelandais, L. (2025). *Real-time pollutant identification through optical PM micro-sensor.* arXiv preprint arXiv2503.10724. https//doi.org/10.48550/arXiv.2503.10724

Choi, Y., Lee, H. J., & Spengler, J. D. (2021). Ambient and personal exposure to air pollutants and the role of microenvironments A review on sensor-based studies. *Science of The Total Environment, 766*, 144303. https//doi.org/10.1016/j.scitotenv.2020.144303

Clark, L. P., Zilber, D., Schmitt, C., Fargo, D. C., Reif, D. M., Motsinger-Reif, A. A., & Messier, K. P. (2025). A review of geospatial exposure models and approaches for health data integration in environmental health applications. *Journal of Exposure Science & Environmental Epidemiology, 35*, 131–148. https//doi.org/10.1038/s41370-024-00712-8

Divino Pouso Mariano, K., Leite Nakano Almada, F., & Adriell Dutra, M. (2024). *Smart air quality monitoring for automotive workshop environments.* arXiv preprint arXiv2410.03986. https//doi.org/10.48550/arXiv.2410.03986arXiv

Lee, S., & Kim, H. (2022). Wearable environmental sensors for real-time personal exposure monitoring A review of recent developments. *Sensors, 22*(9), 3325. https//doi.org/10.3390/s22093325

Li, Z., Khananian, A., Fraser, R. H., & Cihlar, J. (2001). Automatic detection of fire smoke using artificial neural networks and threshold approaches applied to AVHRR imagery. *IEEE Transactions on Geoscience and Remote Sensing, 39*(3), 803–810. https//doi.org/10.1109/36.917906

Ma, X., Liu, R., & Zhan, Y. (2024). ML-driven analysis of ozone exposure in China. *JGR Atmospheres, 129*(5), e2024JD041593. https//doi.org/10.1029/2024JD041593

Mallik, M., Clavier, L., & Gaillot, D. P. (2025). *Comparative analysis of realistic EMF exposure estimation from low-density sensor network by finite & infinite neural networks.* arXiv preprint arXiv2504.07990. https//doi.org/10.48550/arXiv.2504.07990

Mariano, K. D. P., Almada, F. L. N., & Dutra, M. A. (2024). *Smart air quality monitoring for automotive workshop environments.* arXiv preprint arXiv2410.03986. https//doi.org/10.48550/arXiv.2410.03986

Nieuwenhuijsen, M. J. (2020). Influence of sensor technologies and big data on occupational and environmental health research. *Annals of Work Exposures and Health, 64*(3), 297–300. https//doi.org/10.1093/annweh/wxaa010

Patton, A. N., Medvedovsky, K., Zuidema, C., Peters, T. M., & Koehler, K. (2022). Probabilistic machine learning with low-cost sensor networks for occupational exposure assessment and industrial hygiene decision making. *Annals of Work Exposures and Health, 66*(5), 580–590. https//doi.org/10.1093/annweh/wxab105

Shahnavaz, F., & Akhavian, R. (2021). *Automated estimation of construction equipment emission using inertial sensors and ML models.* arXiv preprint arXiv2109.13375. https//doi.org/10.48550/arXiv.2109.13375

Snyder, E. G., Watkins, T. H., Solomon, P. A., Thoma, E. D., Williams, R. W., Hagler, G. S. W., ... & Preuss, P. W. (2013). The changing paradigm of air pollution monitoring. *Environmental Science & Technology, 47*(20), 11369–11377. https//doi.org/10.1021/es4022602

Wei, J., Li, Z., Lyapustin, A., Wang, J., & Dubovik, O. (2023). First close insight into global daily gapless 1 km PM2.5 pollution, variability, and health impact. *Nature Communications, 14*, 7654. https//doi.org/10.1038/s41467-023-43409-9

Wei, J., Wang, J., Li, Z., Kondragunta, S., & Anenberg, S. (2023). Long-term mortality burden trends attributed to black carbon and PM2.5 from wildfire emissions across the continental USA from 2000 to 2020 A deep learning modelling study. *The Lancet Planetary Health, 7*(12), e930–e939. https//doi.org/10.1016/S2542-5196(23)00230-7

Yang, H., Li, L., & Gao, J. (2019). A deep learning approach to exposure estimation using wearable sensors and context information. *Sensors, 19*(23), 5120. https//doi.org/10.3390/s19235120

Yu, C. (2022). Wearable sensor platforms for real-time air quality monitoring in occupational settings. *Sensors, 22*(9), 3500.

Zhou, Y., Han, M., & Wang, J. (2023). Interpretability in machine learning models for occupational health risk assessment. *Journal of Occupational Health Psychology, 28*(2), 135–148. https//doi.org/10.1037/ocp0000321