DOI 10.5162/EUROSENSORS2025/T4.2.1

Short-Term Pollution Prediction Using Personal Environmental Monitoring and Machine Learning

Feiling Pan[1], J.A. Covington[1]
[1] School of Engineering, University of Warwick, Coventry, CV4 7AL, UK
Feiling.Pan@warwick.ac.uk

Summary:

Due to increasing global concerns around air pollution, research on portable environmental monitoring has become an expanding area of research. This study used an in-house developed personal environment monitor, coupled to machine learning, to identify scene switch recognition and short-term pollution prediction. Here, decision trees excelled in environmental identification (97% accuracy), while XGBoost performed best in pollution prediction (R^2 = 0.93). This study provides an effective scheme for short-term pollution warnings, which in the future could be used to warn users of potential harm.

Keywords: Personal environment monitoring, machine learning, short-term pollution prediction, environmental identification, Wearable device

Introduction

As environmental challenges become increasingly complex, personal environmental monitoring equipment has emerged as an important area of research. Unlike traditional environmental monitoring methods, these personal devices offer a significant advantage - the ability to provide precise, real-time ambient data about an individual's immediate environment. Current research on portable monitoring devices covers a range of temporal predictions, from hourly short-term forecasts to longer-term trend analyses. However, the existing prediction models face challenges, particularly when dealing with rapidly changing environments and quick scenario transitions. The current prediction ranges often fall short of meeting the risk assessment needs of personal monitoring devices. This limitation is especially pronounced in short-term prediction research, where the ability to anticipate environmental changes within brief time frames remains a significant challenge [1].

To address this issue, we have divided this task into two stages. First, we attempt to identify a change in the user's environment, which could lead to new pollution risks. This includes identifying the likely location (e.g. outside, coffee shop etc.) the user is in. With this information, we then use machine learning to predict future pollution levels based on the changing levels in that location, therefore providing a targeted warning to the user. We believe that this approach will allow users to make decisions before being exposed to high levels of pollution and make an intervention, such as walking in a different direction.

Material and Methods

This study used an internally developed personal sensing device [2], which can monitor 10 environmental pollution indicators, such as temperature and humidity, CO_2, light intensity, ultraviolet, volatile organic compounds (VOC), NO_X and noise. Using this multi-dimensional environmental data, it uses an in-house developed scoring system to provide an overall environmental score. Fig.1 shows the design of the device and its accompanying APP.

Fig. 1. Design of Monitor unit and App

The environmental score (S), range 0 (risk) to 100 (comfort), is determined based on an assessment of the ten environmental parameters. For each parameter reading, it considers whether it is within a comfort range, discomfort range, or risk range. These individual sensor values are then combined to create a final "environmental score", which is displayed to the user.

This study conducted field testing in the city of Coventry, UK, using the device described above, covering various scenarios and a range of different environments. 1,178 environmental data points were collected and annotated for the specific environment. Using this data, machine learning models were created and compared in their performance in short-term pollution prediction. These models included Long Short-Term Memory (LSTM), Extreme Gradient Boosting (XGBoost), and Convolutional Neural Networks (CNN). The study employed a sliding window mechanism and set time steps to achieve environmental risk prediction approximately 30 seconds in advance.

Result and Discussion

Fig. 1 shows some environmental score data for different locations within Coventry, including restaurants, bus, train station, coffee shops, etc. Tab. 1 shows the results of the machine learning models used to identify the location of the user.

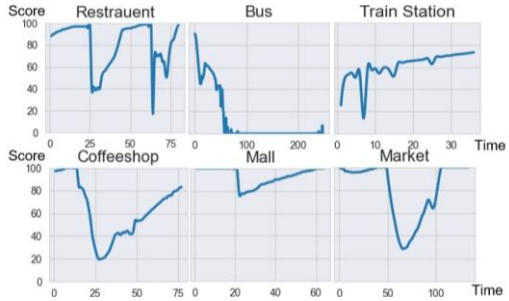

Fig.1. Score changes for typical environments

Tab.1 summarises the performance of each model in identifying the type of environment.

Tab. 1: Environmental identification performance

Model	Accuracy
Decision Tree	97.03%
Random Forest	96.61%
SVM	93.22%
Naive Bayes	89.83%

As can be seen from the table, the decision tree and random forest have the highest accuracy in environmental scene classification.

In addition, this study used a variety of models for short-term (30 seconds in advance) pollution score prediction. Score prediction combines data from multiple pollutants into a simple single predictive value to make it easier for the user. This not only adapts to changes in different environments, but also improves the accuracy of forecasting, helping users make faster decisions.Tab.2 shows some results from three models.

Tab. 2: Short-term pollution warning performance

Model	R^2	MAE	RMSE
XGBoost	0.93	4.96	10.66
CNN	0.91	5.79	12.02
LSTM	0.89	7.89	13.57

According to Tab.2, XGBoost performs best, followed by CNN and LSTM. Fig.2(a), (b) and (c) respectively show the training performance of models. XGBoost performs best at contamination risk (0 value), with concentrated residuals, minimal errors, and stable predictions. CNN is second, there is still some deviation in the low pollution risk area, but the overall stability.

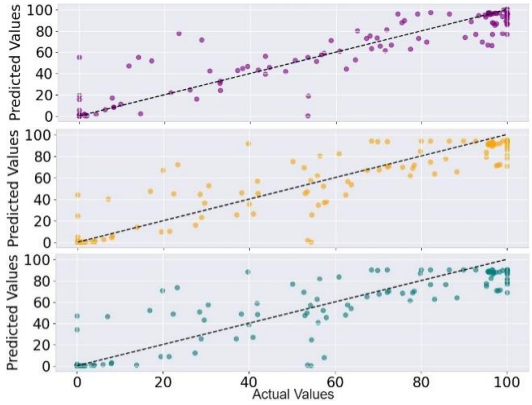

Fig.2. Prediction Accuracy Scatter Plot for Environmental Score: XGBoost (purple), CNN (orange), LSTM (teal)

Conclusions

This study demonstrates the effect of improving short-term pollution prediction and environmental identification through personal environmental sensor data. In terms of environment recognition, the decision tree and random forest model perform well. For short-term pollution prediction, XGBoost has the best effect, followed by CNN and LSTM, which can effectively capture the nonlinear change of environmental pollution. In the future, optimised models can provide more accurate warning and personalised environmental assessment. This may, for example, allow users to adjust their trips, making such environmental predictions closer to their needs.

References

[1] Feng, Y., Liu, S., Wang, J., Yang, J., Jao, Y. L., & Wang, N., Data-driven personal thermal comfort prediction: A literature review, *Renewable and Sustainable Energy Reviews*, 161 (2022) 112357.doi: 10.1016/j.rser.2022.112357

[2] F. Pan, J. A. Covington, A Portable Personalized Environmental Quality Monitoring System (PONG) Version 3, *IEEE Sensors Journal* (2024); doi: 10.1109/JSEN.2024.3405858