Modelling Air Quality via Machine Learning and IoT Technologies

Tushar Saini

(S19011)

Supervisor: Dr. Varun Dutt

School of Computing and Electrical Engineering
Indian Institute of Technology Mandi

This dissertation is submitted for the degree of Master of Science (by Research)

Jan 2022

Thesis Certificate

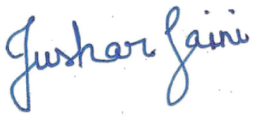
This is to certify that the work contained in the thesis entitled **"Modelling air quality via machine learning and IoT technologies"** being submitted by **Mr Tushar Saini** (Enrolment No: S19011) has been carried out under my supervision. In my opinion, the thesis has reached the standard fulfilling the requirement of regulation of the MS degree. The results embodied in this thesis have not been submitted elsewhere for the award of any degree or diploma.

03rd Jan 2022

Dr. Varun Dutt (Supervisor)
Associate Professor
School of Computing and Electrical Engineering
School of Humanities and Social Sciences
Indian Institute of Technology Mandi Kamand,
Himachal Pradesh, India
Email: varun@iitmandi.ac.in

Declaration by the Research Scholar

I hereby declare that the entire work embodied in this Thesis is the result of investigations carried out by me in the ***School of Computing and Electrical Engineering***, Indian Institute of Technology Mandi, under the supervision of **Dr. Varun Dutt**, and that it has not been submitted elsewhere for any degree or diploma. In keeping with the general practice, due acknowledgments have been made wherever the work described is based on findings of other investigators.



Place: IIT Mandi, Kamand

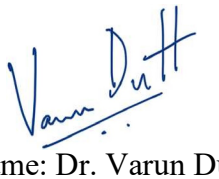
Date: 03rd Jan 2022

Name: Tushar Saini

Declaration by the Research Advisor

This is to certify that the Thesis entitled "**Modelling air quality via machine learning and IoT technologies",** submitted by **Tushar Saini** to the Indian Institute of Technology Mandi for the award of the Degree of Master of Science (by research) is a bonafide record of research work carried out by him under my supervision. The content of this Thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any Degree or Diploma.

Place: IIT Mandi, Kamand



Date: 03rd Jan 2022 Name: Dr. Varun Dutt

Acknowledgements

It is with immense gratitude that I acknowledge the support of my supervisor, Dr. Varun Dutt, in completing this thesis. I am immensely thankful for his invaluable guidance, constructive feedback, and constant encouragement throughout my MS that I was able to complete the research work. The experience which I have obtained has significantly influenced the way I understand and approach research work. I also want to extend my gratitude towards my Annual Progress Committee members, Dr. A. D. Dileep, and Dr. Shyamashree Dasgupta, for their time and guidance throughout my MS. I would also like to acknowledge Dr. Samar Agnihotri for their feedback on my work.

This research work was made possible by the grant provided by the Department of Environment Science and Technology, Government of Himachal Pradesh, on the project IITM/DST-HP/VD/240 to Dr. Varun Dutt and Er. Pratik Chaturvedi. We are also grateful to the Indian Institute of Technology, Mandi, HP, India, for providing computational support, which made training models faster and efficient.

I am also thankful to my fellow lab members Praveen Kumar, Ankush Pathania, Gitanshu Choudhury, Harsh Katakwar, Shashank Uttrani, and Shashank Bhargav, and my friend Punit Pankaj Dubey for their invaluable support and guidance during my short but fulfilling time here on the campus.

Finally, I am extremely grateful to my parents Er. Surendra Singh Saini, Mrs. Sarita Saini, and my brother Piyush Saini for their unconditional support and encouragement throughout my life.

Abstract

Air quality is degrading in developing countries, and severe health issues are associated with increasing air pollution. It is imperative to monitor and predict air quality online in real-time. Although offline air-quality monitoring using hand-held devices is common, online air-quality monitoring is still expensive and uncommon, especially in developing countries. Also, existing technologies do not perform forecasting of air pollution ahead of time. This thesis's first objective was to address these literature gaps and propose a scalable, low-cost real-time air-quality monitoring, prediction, and warning system (AQMPWS). The proposed AQMPWS monitored and predicted seven pollutants: $PM_{1.0}$, $PM_{2.5}$, PM_{10}, carbon monoxide, nitrogen dioxide, ozone, and sulphur dioxide. The AQMPWS also monitored five weather variables: temperature, pressure, relative humidity, wind speed, and wind direction.

This thesis' second objective involved the forecasting of air pollution ahead of time. Individual and ensemble univariate and multivariate time-series forecasting models were developed and tested for predicting air pollution, which could forecast one step ahead in time (i.e., forecasting was done on the time dimension). Five individual time-series forecasting models, namely, multilayer perceptron (MLP), convolution neural network (CNN), long short-term memory (LSTM), and seasonal autoregressive integrated moving average (SARIMA), were investigated. Also, a new weighted ensemble model of these individual models was developed. Among the individual univariate models, the CNN performed the best, and this model was followed by the LSTM, MLP, and SARIMA models both during training and test. The investigation revealed that multivariate models performed better than their counterpart univariate models. Furthermore, the weighted ensemble model performed the best among all models.

In the third objective, the developed AQMPWS was benchmarked against an industrial-grade air-quality monitoring system deployed at the ACC Cement Factory in Barmana, India. A high correlation between the collected data from AQMPWS and the industrial-grade system showed that the developed system was on par with the industrial system. Three statistical models, namely, Vector Autoregressive (VAR), Vector Autoregressive Moving Average (VARMA), and SARIMA and an ensemble model, were developed to predict the future values of $PM_{2.5}$ and PM_{10} in the industrial system from the $PM_{2.5}$ and PM_{10} values measured by the

AQMPWS. Results showed that the ensemble model performed the best, and VAR performed the second-best in predicting the $PM_{2.5}$ and PM_{10} values in the industrial system.

Lastly, we performed a preliminary study to find the public perception of the adverse effects of air pollution. About 90 percent of those surveyed (N = 500) requested a daily update on their localities' air pollution. Results showed that people preferred periodic SMS alerts to learn about air quality compared to other methods. Results from the study helped shape the alerting system in the AQMPWS.

Keywords: *Air pollution, IoT, Time-series forecasting, Machine Learning, Deep Learning, Statistical Modelling*

Table of Contents