APPLICATION OF AI AND BIG DATA TO AIR POLLUTION MONITORING AND FORECASTING:

Experiences of some countries around the world and proposed solutions for Vietnam

TRẦN HỮU SỸ
Vietnam Academy of Science and Technology

Air pollution (AP) is currently considered one of the most serious threats to public health and the path to realizing the global Sustainable Development Goals (SDGs). Faced with this reality, more and more countries around the world are prioritizing the choice of modern technology, considering this as the main solution in the fight against AP, in which, artificial intelligence (AI) and big data (Big Data) are emerging as breakthrough tools, helping to analyze complex environmental data from many different sources. Thanks to their fast and accurate processing, these systems can identify pollution "hot spots" in real time; simulating the trend of dispersing hazardous substances, and at the same time providing a scientific foundation to develop effective and timely response policies. The article introduces a number of typical countries in the world that have successfully applied AI and Big Data to AP monitoring and forecasting, thereby proposing an appropriate direction for Vietnam.

1. SOME TYPICAL COUNTRIES HAVE SUCCESSFULLY APPLIED AI AND BIG DATA TO AIR POLLUTION MONITORING AND FORECASTING

In recent years, AI and Big Data have emerged as a widely used technological tool to control and minimize the harm of toxic substances that cause ethanol [7]. The development of the internet in the 21st century has allowed AI to reach its full potential and apply this technology to model complex environmental problems, especially air quality (AQ) regulation. To diagnose, monitor, and cure a number of diseases related to AP, many researchers around the world have used AI approaches in healthcare decision-making tools [8]. For example, Heuvelmans et al., have developed a deep learning-based approach to simulate the progression of cancer cells in the lungs using CT scan image datasets [9]. Polezer et al., again, used artificial neural networks (Multilayer Perceptron (MLP), Echo State Networks (ESN), and Extreme Learning Machines (ELM) to assess the negative effects of atmospheric AP on individual health [10].

This article mentions the US, China, Singapore... are proofs that the application of advanced technology in AP monitoring can bring positive results in both accuracy and coverage. Notably, more and more people around the world are actively using AI-integrated devices and applications to monitor AQ. This not only raises public awareness of environmental issues, but also plays an important role in supplementing practical data, contributing to the improvement of effective pollution warning and management systems.

1.1. The United States - A pioneer in applying AI and Big Data to air quality management

The US has built many modern monitoring systems that effectively support urban pollution control, including the AirNow System operated by the US Environmental Protection Agency (EPA). This system collects data from more than 4,000 monitoring stations nationwide, combines AI algorithms to analyze and provide real-time AQ information. Each year, more than 100 million people in the U.S. access AirNow through websites, mobile apps, and public electronic boards. According to statistics from the EPA, thanks to the application of this technology, the number of days with high pollution levels in big cities has decreased by 8 - 10% in just 5 years [1].

In addition, the TEMPO Sensor System, a spectrophotometer mounted on a geostationary orbit satellite, developed by the US National Aeronautics and Space Administration (NASA), capable of tracking AQ variants on a suburban scale, marks an important step forward in large-scale AP monitoring. This device allows the data collection of several contaminants such as nitrogen dioxide (NO_2), ozone (O_3), sulfur dioxide (SO_2) and formaldehyde (CH_2O) with high resolution, hourly frequency. As a result, scientists can continuously monitor changes in the AQ and trace emissions in urban areas along the stretch from Canada to Mexico. TEMPO is a clear demonstration that AI and Big Data are being applied more and more extensively in an effort to control AP.

In addition, many major cities in the US such as New York, Los Angeles, Chicago... are also deploying mobile sensor systems mounted on buses, drones, and public bicycles to enhance air monitoring capabilities at the micro level. According to the Massachusetts Institute of Technology (MIT) Senseable City Lab, these sensors can detect localized pollution points smaller than 100 meters - areas that are often missed by traditional fixed surveillance systems. In particular, air data coverage in the inner city area has increased from less than 40% to more than 85% [1]. On the other hand, from 2021, AI will be used in smart grid systems, helping to optimize energy consumption, reducing greenhouse gas emissions by 15 - 20% in US states.

1.2. China: Monitoring and controlling air pollution through AI and Big Data China is focusing on investing in the application of AI and Big Data to manage environmental protection, through the deployment of the "Skynet" system of more than 20 million surveillance cameras, combined with data from environmental sensors to monitor and control environmental and water pollution. At the same time, the country has also implemented strong AI in its monitoring program, with a notable step being the automation of environmental monitoring stations, a combination of AI, drone equipment, and automated laboratories [1]. In the field of monitoring water and water, the upgrading and digitization of national monitoring stations will help automate the monitoring process without the need for operating personnel, thereby reducing the frequency of maintenance and processing time by more than 70%. Sampling and analysis are also made easier with drone sampling technology and automatic laboratory sample analysis, helping to save more than 50% of time, manpower, and transportation costs [2]. In addition, China also uses AI to predict dense fog, thereby providing early warning measures for citizens.

Currently, many major Chinese cities such as Beijing, Shanghai, Shenzhen, and Guangzhou have implemented smart city models, AI applications, the internet of things (IoT), Big Data and environmental sensors to monitor AQ, pollution control, power consumption forecasting and waste management, etc wastewater. Digital technology also supports optimal traffic regulation, public lighting, early warning of the environment and improving the efficiency of urban governance. A 2022 study by Tsinghua University showed that the application of AI helped reduce the level of COVID-19 in Beijing by 12% during the winter. China is also promoting the development of green e-commerce, digital finance for green projects, blockchain application in traceability, and agricultural supply chain transparency. The models of "digital carbon credit" and "digital carbon bank" have been tested in many localities, helping businesses and people participate in the carbon market more transparently and effectively. In particular, China is a pioneer in integrating digital solutions into green manufacturing, smart agriculture and renewable energy. Many smart factories apply AI, robots, and IoT to help optimize production processes, reduce resource waste and greenhouse gas emissions. In agriculture, automatic irrigation systems, environmental sensors, and product traceability contribute to saving resources, limiting emissions, improving product quality, and achieving sustainable development goals [3].

Not only that, China is also known as a leading country in monitoring biodiversity and environmental noise with smart sensor equipment, thereby being able to detect ecological abnormalities early and improve the accuracy of monitoring. Smart devices such as infrared cameras, bird voice recorders, amphibian and reptile monitoring radars, butterfly monitoring devices, etc. have helped automate the monitoring process, achieving an accuracy of over 85%. In the future, China's Ministry of Ecology and Environment will continue to promote the application of digital technology to manage discharge points according to the "one point per code" model, in order to ensure that data can be connected, retrieved and monitored closely. Some technologies such as satellite remote sensing, automatic monitoring will also be applied to detect and promptly handle illegal discharges that cause environmental pollution [2]. These achievements demonstrate the central role of digital technology in green economic development, contributing to pollution control, emission reduction, and improving the quality of life, while strengthening China's international integration capacity.

1.3. Singapore: Technology is the foundation for environmental strategy and smart cities

As a small country with a high population density and commercial activity, Singapore is under great pressure to control the AP, so from a very early age, the Singapore Government has been deeply aware of the key role of digital technology and AI in the future of the country's development; identifying this as an important foundation for green urban development, helping the city effectively adapt to all environmental challenges, towards the goal of sustainable development.

As early as the 2010s, Singapore has begun to implement the vision of a "smart nation", integrating Big Data, AI, IoT into all socio-economic fields to improve the quality of life of its citizens. One of the typical initiatives is the MAREMIS Project, in cooperation with the Port of Hamburg (Germany), which uses AI to monitor and regulate emissions from ships. The system analyzes transportation data, wind speed, and humidity to predict NO_x and SO_x emissions, thereby supporting accurate decision-making. According to the Singapore National Environment Agency (NEA), MAREMIS has helped reduce NO_x emissions in the port area by an average of 18% after 3 years of implementation [1]. In addition, the "OneMap", "Smart Water" projects

The American AirNow system collects data from more than 4.000 monitoring stations nationwide, combining AI algorithms to analyze and provide real-time AQ information

and environmental sensor systems help monitor and forecast water and water conditions; optimizing water supply and drainage operations, contributing to saving resources and preventing flooding. Lighting systems, public transportation are automated, reducing energy consumption, greenhouse gas emissions and promoting green transportation. Smart residential areas also connect data between electricity, water, transportation, and health to optimize operation and control emissions at the source.

Singapore is also a country that stands out for its comprehensive smart management solutions for infrastructure and public services, such as environmentally friendly air, water, lighting, waste and transportation control. The "Virtual Singapore" model - 3D Virtual City is a revolutionary creation. The system integrates a wide range of data sources from environmental sensors, traffic systems, security cameras, and weather maps. As a result, managers can simulate policies, test solutions, and predict effectiveness before implementing them in practice. According to the NEA, the use of Virtual Singapore in the simulation of inner-city traffic adjustments has helped reduce CO_2 emissions from private vehicles by more than 10% after just two years of testing in the Tengah district. GS. Michael Batty, University College London commented, "Virtual Singapore is a prime example of the city of the future, where every decision on environmental planning and infrastructure is based on real-time data and advanced predictive modeling" [1].

2. PROPOSE SOME SOLUTIONS FOR VIETNAM

2.1. Advances in the application of AI and Big Data to air pollution monitoring and forecasting

Fi-Mi: A mobile system for monitoring and predicting air quality with AI application

In Vietnam, in recent years, the application of AI and Big Data in the field of environment and pollution reduction has made significant progress. Specifically, in order to increase the accuracy of low-cost monitoring devices and minimize the number of devices used, a Vietnamese research team from Hanoi University of Science and Technology, Chiba University (Japan) and Toshiba Corporation has successfully coordinated to implement the project "Fi-Mi: Mobile system for monitoring and predicting AQ with AI application" under the sponsorship of Vingroup Innovation Fund (VinIF). With the Fi-Mi system, the research team has built a system to monitor AP parameters and emission sources affecting human health, contributing to solving two main problems: (i) Automatically calibrate the data collected by low-cost monitoring equipment, in order to increase accuracy; (ii) Prediction of monitoring data in space and time, in order to build a high-resolution AQ map, while not needing to deploy too many instruments [4].

The Fi-Mi system consists of three layers (sensors, information, applications), in which, the sensor layer consists of air monitoring devices, which are placed on vehicles (buses), continuously collect AQ indicators and send them to the server system. The information layer consists of communication protocols, which take care of the transmission and receipt of data between the sensor layer and the application layer. The application layer includes a server system, which is responsible for processing, storing, and displaying visualized data for users; use machine learning models to predict AQ in space and time. With a compact structure and low price, Fi-Mi monitoring devices are suitable for installation on buses and cars moving around the city, helping to collect temperature, humidity, $PM_{2.5}$, SO_2, NO_2, CO indicators. real-time data on the server system. In addition to information about air and weather monitoring, the devices also send to the server signal strength/weakness information (RSSI) and the location of the device collected through the GPS module. The entire operation of each Fi-Mi monitoring device is controlled by the MCU central controller programmed by the research team at Hanoi University of Science and Technology [5].

Not only stopping at the problem of predicting AQ over time, one of the outstanding applications of Fi-Mi is to study methods to estimate the distribution of AQ, that is, using AQ data obtained from the locations of monitoring equipment to estimate AQ in places where there is no monitoring equipment. On the other hand, the Fi-Mi Project builds a deep learning model, using graph neural networks to represent the spatial relationship between observation stations. This is also the first project to use deep learning in PM2.5 prediction, implemented on Vietnam's dataset. The results of the evaluation on the AQ dataset collected in Hanoi showed that the Encoder-Decoder model reduced the prediction error by 53.7% compared to the solution of some other studies. Furthermore, the use of an input feature auto-selection algorithm can help reduce prediction error by 13.7% compared to other algorithms. In addition, the combination of meteorological information, especially wind, to increase the accuracy of predictions is also a strength of Fi-Mi. Results from several experimental studies show that Fi-Mi's solution has helped reduce the prediction error from 4.93% to 34.88% compared to current methods [6].

After the project ends and is accepted in 2023, the team of scientists at Ha Noi University of Science and Technology together with researchers in the field of AI and the environment continue to pursue related research, aiming to build a complete system, which will not only help track AP parameters but also the source of emissions as well as the impact of AP on human health.

Application of Big Data in remote sensing to monitor and reduce air quality from waste treatment sites

From July 2020 to December 2022, scientists from the National Remote Sensing Agency, Ministry of Agriculture and Environment implemented the project "Research on the application of Big Data - Remote sensing in monitoring AP from waste treatment (WT) areas" through the use of data from the Sentinel-5P satellite and ground monitoring to build AQ maps; determine the dispersal of some pollutants such as SO_2, NO_2, CH_4, etc. As a result, the research team has built a scientific basis and methodology for the application of Big Data - Remote sensing to monitor AP from waste treatment zones; the application combines specialized remote sensing data Sentinel-5P, specialized UAV flight data and ground observation data for regular and high-granular monitoring of the AQ around the WT area. At the same time, it is possible to build a technological process for applying Big Data - Remote sensing to monitor the status of AP from waste treatment sites; propose steps, from remote sensing data processing, remote sensing and monitoring data integration, and modeling of AQ propagation; identified the spread of a number of toxic chemical components that cause NOKK from concentrated wastewater treatment areas such as SO_2, NO_2, CH_4. In addition, the research team collected and processed Sentinel-2 remote sensing data; Sentinel-5P, Landsat-8/9; data collected from UAVs; monitoring data to calculate a number of AQ components such as $PM_{2.5}$, PM_{10}, CO, SO_2, NO_2, CH_4; collect ozone data as a basis for building the AQI24h map.

After publication, the research results are transferred to the agencies tasked with monitoring and supervising the waste management in general, the pollution caused by the operation of concentrated landfills in particular by the method of direct transfer accompanied by training and user manuals.

The system of monitoring, forecasting and warning of air quality on the basis of collecting and integrating multi-source data

In order to build a real-time monitoring, forecasting and warning system for AQ applications of IoT technology, cloud computing (Cloud), AI and numerical modeling, Assoc. Prof. TS. Pham Tran Vu, University of Science and Technology, Vietnam National University of Ho Chi Minh City and his colleagues have successfully implemented the project "Building a system of monitoring, forecasting and warning of AQ on the basis of collecting and integrating multi-source data, piloting for a large city", under the National Key Program for the period to 2025 "Supporting research, development and application of technology of industry 4.0", code KC-4.0/19-25.

Implementing the project, the research team integrated AI with the most advanced 3-layer model and machine learning to forecast short-term AP for Ho Chi Minh City in 24 hours, helping to provide early warning of AP to protect people's health as well as reduce economic losses. Specifically, the research team has studied the real-time monitoring, forecasting and warning system of AQ in the world and Vietnam, thereby building wireless communication sensor clusters; the AQ estimation model from satellite imagery and auxiliary data for the Ho Chi Minh City area; the AQ forecasting model using AI; monitoring network, simulation model system for forecasting and warning of AQ on the mobile application platform and website. As a result, a model of real-time air environment monitoring stations has been built connected according to IoT standards: SO_2, CO, NO_2, NO, O_3, PM_{10}, $PM2_{.5}$; monitoring and forecasting system for AQ for pilot application in Ho Chi Minh City. At the same time, the research team has completed the design of the sensor node, serving data collection activities for testing and evaluation; testing AI models and CMAQ models to forecast and warn AQ; completed the model of estimating $PM_{2.5}$ dust concentration from remote sensing images.

The advantage of the system is that when installing the application or accessing the website, people can know the current AQ in an area, and at the same time, be forecasted for the next 1 - 2 days to promptly have a plan to protect their health against the impact of AP. In particular, the application will send a direct message to users when AP shows signs of increase, thereby making appropriate recommendations depending on the level of pollution.

2.2. Challenges, barriers and solution proposals

Challenges and barriers

Despite the initial success, Vietnam still faces many barriers in the application of AI and Big Data in forecasting and warning of environmental pollution in general and AP in particular, specifically: (i) Lack of synchronous and high-quality data platform: Environmental data in Vietnam is still scattered, lack of standardization, difficult to integrate, and not ready for AI analysis. Many localities still collect data manually, lack sensor equipment, and do not have an effective sharing mechanism between levels and sectors; (ii) Limited analysis and implementation capacity: The number of AI and data science experts in Vietnam is still low; current AI models are only small experiments, and there is no ecosystem strong enough to be applied on a large scale. The ability to apply AI to public decision-making is still limited, due to the lack of coordination between engineering and management; (iii) Mechanisms and policies have not really created encouragement: Investment in AI and Big Data in the field of environment has not been considered a national priority. Vietnam does not have a clear legal mechanism for data use, privacy protection, information security as well as effective budget allocation for digital technology in the natural resources and environment sector; (iv) Lack of public-private partnership (PPP): The participation of businesses, startups and research institutes in this field is still fragmented, and there is no strong enough incentive mechanism to promote public-private cooperation. Technology projects are often limited by budget or do not have the right testing corridors; (v) Lack of connection with international trends: Vietnam has not participated deeply in global initiatives on environmental technology such as the global atmospheric data network, the AI platform for the environment of the United Nations Environment Programme, or open projects on Earth data, etc. This deprives us of access to advanced technology and large-scale data.

Solution-oriented in the coming time

Firstly, Vietnam needs to establish a comprehensive national environmental data system, connecting data from monitoring stations, IoT sensors, satellite imagery and local reports. Data must be standardized, expanding the ability to be shared and used by state agencies, research institutes, and businesses. This requires the Government to play a central role in the design of the legal framework and technical infrastructure.

Secondly, the development of AI testing models in the field of AP monitoring in large cities in the form of PPP should be encouraged to take advantage of the strengths of the private sector. At the same time, AI and Big Data need to be integrated into decision-making support tools in developing pollution forecasting policies for urban planning, environmental impact analysis (EIA), etc. On the other hand, it is necessary to soon include contents related to the application of AI and Big Data in monitoring and forecasting in undergraduate and postgraduate training programs, professional training to increase the number as well as improve the capacity of a team of experts who are strong enough in data science. AI and environmental management.

Thirdly, it is possible to participate in international programs related to the development and implementation of environmental policies, especially the reduction of AP, such as Copernicus - the European Union's Earth Observation Program, which provides free satellite data on the environment, thereby supporting member states in monitoring and managing natural resources; initiatives from the United Nations Environment Program, the World Bank on environmental technology, etc. in order to learn from experience, access to global data, open source technology and financial and technical support to implement a number of pilot projects with high pervasiveness.

Fourthly, it is necessary to soon complete the National Action Plan to overcome pollution and manage pollution, with specific goals for 2025 and the period 2026 - 2030, focusing on main groups of solutions such as energy, waste sources, transportation, and construction; strengthen the application of AI and Big Data to analyze and forecast AQ; integrate automatic monitoring systems and national databases, helping localities periodically inventory emissions and publicize information transparently.

Fifthly, AI and Big Data bring transformative power in efforts to solve the problem of urban climate change, this tool helps improve the capacity to measure, forecast, optimize and act, thereby supporting the construction of smarter, cleaner, fairer and more livable cities. In order for digital technology to truly be a force for the common good, it is necessary to ensure 3 criteria: (i) Inclusivity, in which, all people can access benefits, not only those who own smartphones or wearable devices; (ii) Equity (The system must be designed not to transfer pollution to communities that are already environmentally burdened); (iii) Transparency and ethics (Community trust requires clear governance, open access, and substantive citizen participation). Therefore, AI and Big Data should not be something that is "imposed on humans", but must be tools built "with humans". That means opening up access to data, encouraging open-source development, and incorporating local voices into the design and deployment of tools. At the same time, AI and Big Data must be key tools if we want to achieve global goals - as announced by the World Health Organization at the Global Conference on AP and health, to reduce the health impact of AP by 50% by 2040. To realize the above goal, Vietnam cannot rely on algorithms alone, but needs intentional coordination between the Government, researchers, startups, urban planners and the whole community.

3. CONCLUSION

AI and Big Data are being applied more and more extensively in an effort to control AP. Practical experiences from the United States, China, and Singapore show that technology is no longer just a technical tool but has become a solid foundation in environmental and health policymaking. Like many countries around the world, with the strong development of AI and Big Data, AP monitoring and forecasting in Vietnam is becoming more accurate, faster and more effective, helping to minimize negative impacts on the environment as well as human health, towards the goal of sustainable development■

REFERENCES

1. Tung Lam, Long Hai, 2025. Application of AI and big data in ONKK monitoring and forecasting. Try to update at https://kinhtedothi.vn/ung-dung-ai-va-du-lieu-lon-trong-giam-sat-va-du-bao-o-nhiem-khong-khi.671326.html.

2. Ministry of Ecology and Environment of China, 2025. Information from the Director of the Supervision Bureau of the Ministry of Ecology and Environment of China Chiang Huohua at the regular press conference in March, which will take place on March 26, 2025.

3. Wang, C., Liu, T., Du, D., Zhu, Y., Zheng, Z., & Li, H. (2024). Impact of the Digital Economy on the Green Economy: Evidence from China, Sustainability, 16 (21), 9217.

4. Assoc. Prof. TS. Nguyen Phi Le, Hanoi University of Science and Technology, 2025. Application of AI in air pollution monitoring and forecasting in Vietnam. Intermediate at https://blog.vinbigdata.org/ung-dung-ai-trong-quan-trac-va-du-bao-o-nhiem-khong-khi-tai-viet-nam/#:~:text=C%C3%A2u%20h%E1%BB%8Fi%20n%C3%A0y%20l%C3%A0%20%C4%91%E1%BB%99ng,s%C3%A1ng%20t%E1%BA%A1o%20Vingroup%20(VinIF).

5. Viet An Nguyen, Viet Hung, Van Sang Doan, Thanh Hung Nguyen, Phan Thuan Do, Kien Nguyen, Phi Le Nguyen, Minh Thuy Le, "Realizing Mobile Air Quality Monitoring System: Architectural Concept and Device Prototype", Asia Pacific Conference on Communications (APCC 2021).

6. Nguyen Minh Hieu, Phi Le Nguyen, Kien Nguyen, Thanh Hung Nguyen, and Yusheng Ji. "PM2. 5 prediction using genetic algorithm-based feature selection and encoder-decoder model." IEEE Access 9 (2021): 57338 - 57350.

7. Mo, X.; Zhang, L.; Li, H.; Qu, Z. A new AI-based AQ early warning system. Int. J. Environ. Res. Public Health 2019, 16, 3505.

8. Masood, A.; Ahmad, K. Review of Emerging AI Techniques for ONKK Forecasting: Basics, Applications, and Performance. J. Clean. Prod. 2021, 322, 129072.

9. Heuvelmans, MA; van Ooijen, PM; Ather, S.; Silva, C.F.; Han, D.; Heussel, CP; Hickes, W.; Kauczor, H.-U.; Novotny, P.; Peschl, H. Predicting Lung Cancer Using Deep Learning to Identify Benign Lung Nodules. Lung Cancer 2021, 154, 1 - 4.

10. Polezer, G.; Tadano, Y.S.; Siqueira, HV; Godoi, A. F.; Yamamoto, C.I.; de André, PA; Pauliquevis, T.; de Fatima Andrade, M.; Oliveira, A.; Saldiva, P.H. Assessment of the impact of PM2.5 on respiratory disease using artificial neural networks. Environment. Pollution. 2018, 235, 394 - 403.