**TÓM TẮT THÔNG TIN VỀ LUẬN ÁN**

Tên đề tài luận án: **Nghiên cứu ứng dụng phương pháp học máy và học sâu   
dự báo nồng độ bụi PM2.5 trong không khí xung quanh – Trường hợp   
nghiên cứu cho thành phố Hồ Chí Minh**

Ngành: Quản lý tài nguyên và môi trường

Mã số ngành: 9850101

Họ tên nghiên cứu sinh: Nguyễn Phúc Hiếu

Khóa đào tạo: 2020

Người hướng dẫn khoa học:

1. **PGS.TS. Đào Nguyên Khôi**
2. TS. Nguyễn Lý Sỹ Phú

Cơ sở đào tạo: Trường Đại học Khoa học Tự nhiên - ĐHQG.HCM

1. **TÓM TẮT NỘI DUNG LUẬN ÁN:**

Ứng dụng của các thuật toán học máy và học sâu trong quản lý ô nhiễm không khí ngày càng trở nên quan trọng do khả năng mô phỏng các hiện tượng phi tuyến tính phức tạp và xử lý lượng dữ liệu lớn. Bên cạnh đó, các thuật toán trí tuệ nhân tạo còn giúp mô phỏng và dự đoán ô nhiễm đạt hiệu quả cao và cần ít lượng thông tin đầu vào, phù hợp với những khu vực hạn chế về dữ liệu quan trắc. Ô nhiễm không khí, đặc biệt là ô nhiễm bụi mịn (PM2.5), là một vấn đề thu hút nhiều sự quan tâm tại Thành phố Hồ Chí Minh (TP.HCM) do gây ảnh hưởng nghiêm trọng đến sức khỏe cộng đồng và môi trường.   
Tuy nhiên, vấn đề dự báo cảnh báo vẫn còn nhiều hạn chế.

Trước những thách thức do vấn đề ô nhiễm bụi PM2.5, việc phát triển các phương pháp tiên tiến để dự đoán nồng độ bụi PM2.5 ở TP.HCM là điều cần thiết. Nghiên cứu áp dụng 05 thuật toán học máy gồm Linear Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Regression (SVR), Extreme Gradient Boosting (XGBoost), và 04 thuật toán học sâu gồm General Regression Neural Network (GRNN), Artificial Neural Network (ANN), Convolutional Neural Network (CNN) và Long Short-Term Memory (LSTM) để mô phỏng và dự báo nồng độ bụi PM2.5 trong không khí xung quanh tại   
khu vực trung tâm TP.HCM.

Với bài toán mô phỏng, kết quả nghiên cứu cho thấy ba thuật toán ANN, LSTM và SVR có khả năng mô phỏng nồng độ PM2.5 tốt nhất trong số các thuật toán được lựa chọn. Trong đó, thuật toán ANN đạt hiệu suất mô phỏng tốt nhất với kịch bản 6 có các biến đầu vào gồm độ ẩm, nhiệt độ, tốc độ gió, lượng mưa, độ bốc hơi, số giờ nắng và biến đầu ra là bụi PM2.5.

Với bài toán dự báo, ba thuật toán gồm ANN, SVR, và XGBoost được sử dụng để xây dựng mô hình dự báo ngắn hạn (1, 3, 5 và 7 ngày) cho nồng độ bụi PM2.5. ANN tiếp tục thể hiện ưu thế so với 2 thuật toán còn lại.

Kết quả nghiên cứu có ý nghĩa quan trọng trong quản lý tài nguyên và môi trường. Từ kết quả nghiên cứu, luận án đã xây dựng ứng dụng dự báo nồng độ bụi PM2.5 trong không khí xung quanh tại trung tâm TP.HCM với giao diện trực quan, dễ sử dụng.   
Đồng thời, ứng dụng còn chuyển đổi giá trị nồng độ bụi PM2.5 thành chỉ số chất lượng không khí (AQI) theo quy định của Việt Nam nhanh chóng, và đưa ra mức VN‗AQI tương ứng ảnh hưởng tới sức khỏe, một số hoạt động khuyến nghị giúp người dân có thể chủ động có các hành động phù hợp nhằm bảo vệ sức khỏe.

1. **NHỮNG KẾT QUẢ MỚI CỦA LUẬN ÁN**

* Áp dụng đồng thời 9 thuật toán học máy và học sâu (LR, DT, RF, SVR, XGBoost, GRNN, ANN, CNN, và LSTM) để xác định mô hình tối ưu trong mô phỏng và dự báo nồng độ bụi PM2.5 tại TP.HCM dựa trên các yếu tố khí tượng (nhiệt độ, tốc độ gió, độ ẩm, số giờ nắng, lượng mưa, và độ bốc hơi).
* Kết quả mô phỏng: ANN đạt hiệu suất tốt nhất với kịch bản 6, các chỉ số RMSE = 7,978 µg/m³, MAPE = 32,452%, IOA = 0,736, và NMB = 0,032.
* Kết quả dự báo: ANN là thuật toán tốt nhất cho dự báo sớm 1, 3, 5 và 7 ngày, với chỉ số RMSE = 6,856 µg/m³, MAPE = 29,025%, IOA = 0,765, NMB = 0,019.
* Xây dựng ứng dụng dự báo nồng độ bụi PM2.5 tại TP.HCM được phát triển từ kết quả nghiên cứu của luận án. Ứng dụng cung cấp khả năng dự báo trong nhiều khoảng thời gian khác nhau, bao gồm mô phỏng hiện tại và dự báo sớm 1, 3, 5, và 7 ngày, đáp ứng nhu cầu đa dạng của người sử dụng. Ứng dụng có thể được truy cập thông qua đường dẫn sau: https://hcm-pm25-prediction.streamlit.app/

1. **CÁC ỨNG DỤNG/KHẢ NĂNG ỨNG DỤNG TRONG THỰC TIỄN HAY NHỮNG VẤN ĐỀ CÒN BỎ NGỎ CẦN TIẾP TỤC NGHIÊN CỨU**

* Cải thiện độ chính xác của dự báo: Mở rộng hệ thống quan trắc tại nhiều vị trí khác nhau ở TP.HCM và các khu vực lân cận để tăng độ bao phủ không gian, phản ánh chính xác hơn sự phân bố nồng độ PM2.5.
* Nâng cao mô hình dự báo dài hạn: Tích hợp thêm dữ liệu về giao thông, mật độ dân cư, nguồn thải công nghiệp và yếu tố mùa vụ nhằm đánh giá chi tiết hơn các tác nhân ảnh hưởng đến ô nhiễm không khí, đặc biệt trong các mùa ô nhiễm cao.

|  |  |  |
| --- | --- | --- |
| **TẬP THỂ CÁN BỘ HƯỚNG DẪN** (Ký tên, họ tên) | | **NGHIÊN CỨU SINH** (Ký tên, họ tên) |
| **PGS.TS Đào Nguyên Khôi** | TS. Nguyễn Lý Sỹ Phú | Nguyễn Phúc Hiếu |
| **XÁC NHẬN CỦA CƠ SỞ ĐÀO TẠO KT. HIỆU TRƯỞNG** | | |

**THESIS INFORMATION**

Thesis title: **Research on predicting PM2.5 concentration in the ambient air using machine learning and deep learning algorithms - A case study in Ho Chi Minh city**

Speciality: Natural resources and environmental management

Code: 9850101

Name of PhD Student: Nguyen Phuc Hieu

Academic year: 2020

Supervisor:

1. **Dao Nguyen Khoi (Assoc. Prof. Ph.D.)**
2. Nguyen Ly Sy Phu (Ph.D.)

At: VNUHCM – University of Science

1. **SUMMARY:**

The application of machine learning and deep learning algorithms in air pollution management is increasingly important due to their ability to model complex nonlinear phenomena and handle big datasets. In addition, artificial intelligence algorithms assist in simulating and predicting pollution with high accuracy and require minimal input data. This is suitable for areas with limited access to observational data. Air pollution, particularly fine particulate matter (PM2.5), is a pressing issue in Ho Chi Minh City (HCMC) due to its severe impact on public health and the environment. However, the predicting and warning systems still face significant limitations.

Given the challenges posed by PM2.5 pollution, developing advanced methods to predict PM2.5 concentration in HCMC is essential. This study applies five (05) machine learning including Linear Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Regression (SVR), Extreme Gradient Boosting (XGBoost), and four (04) deep learning algorithms including General Regression Neural Network (GRNN), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) to simulate and predict PM2.5 concentrations in central HCMC.

For the simulation task, the findings reveal that three algorithms, including ANN, LSTM, and SVR, achieved the best performance in modeling PM2.5 concentrations. Among them, ANN performed the best under Scenario 6, which utilizes input variables including humidity, temperature, wind speed, precipitation, evaporation, hours of sunshine, and the output variable of PM2.5.

For the predicting task, three algorithms, including ANN, SVR, and XGBoost, were used to develop short-term predicting models (1, 3, 5, and 7 days ahead) for PM2.5 concentration. ANN continued to show superior performance.

The research results have significant implications for resource and environmental management. From the findings, the research has developed an application for predicting PM2.5 concentrations in the ambient air around the center of HCMC. This application features a user-friendly, intuitive interface. Additionally, it rapidly converts PM2.5 concentration values into the Air Quality Index (AQI) according to Vietnamese regulations, provides the corresponding VN‗AQI levels, and their health implications. It also offers recommended activities that enable residents to take appropriate actions to protect their health.

1. **NOVELTY OF THESIS:**

* This study is to apply nine machine learning and deep learning algorithms (LR, DT, RF, SVR, XGBoost, GRNN, ANN, CNN, and LSTM) to identify the optimal model for PM2.5 concentration modeling and predict in HCMC based on meteorological factors (humidity, temperature, wind speed, evaporation, sunshine hours, and rainfall).
* Best simulating result: ANN performed the best under Scenario 6, achieving RMSE = 7,978 µg/m³, MAPE = 32,452%, IOA = 0,736, and NMB = 0,032.
* Best predicting result: ANN outperformed other models for 1, 3, 5, 7 day predicting, RMSE = 6,856 µg/m³, MAPE = 29,025%, IOA = 0,765 NMB = 0,019.
* The PM2.5 concentration predicting application for HCMC has been developed based on the research results of the research. This application provides predicting capabilities for various time intervals, including current simulations and predict for 1, 3, 5, and 7 days ahead. The application can be accessed through the following link: https://hcm-pm25-prediction.streamlit.app/.

1. **APPLICATIONS/APPLICABILITY/PERSPECTIVE**

* Improving forecast accuracy: Expanding air quality monitoring systems to multiple locations in HCMC and surrounding areas to enhance spatial coverage and better capture PM2.5 distribution patterns.
* Enhancing long-term forecasting models: Integrating additional data sources, such as traffic patterns, population density, industrial emissions, and seasonal factors, to provide a more detailed assessment of air pollution contributors, especially during high-pollution seasons.

|  |  |  |
| --- | --- | --- |
| **SUPERVISOR** | | **PhD STUDENT** |
| **Dao Nguyen Khoi (Assoc. Prof. Ph.D.)** | Nguyen Ly Sy Phu (Ph.D.) | Nguyen Phuc Hieu |
| **CONFIRMATION UNIVERSITY OF SCIENCE**  **VICE PRESIDENT** | | |