

# TỔNG QUAN VỀ ỨNG DỤNG MẠNG NƠRON NHÂN TẠO (ANNs) TRONG DỰ BÁO CHẤT LƯỢNG KHÔNG KHÍ

Nghiêm Trung Dũng<sup>1</sup>, Mạc Duy Hưng<sup>1,2\*</sup>, Hoàng Xuân Cơ<sup>3</sup>

<sup>1</sup>Viện Khoa học và Công nghệ Môi trường - ĐH Bách khoa Hà Nội

<sup>2</sup>Trường Đại học Kỹ thuật Công nghiệp - ĐH Thái Nguyên

<sup>3</sup>Trường Đại học Khoa học Tự nhiên - ĐH Quốc gia Hà Nội

## TÓM TẮT

Mạng nơron nhân tạo (ANN) đã trở thành một hướng tiếp cận hữu hiệu để thay thế các mô hình thống kê truyền thống. Chính vì vậy, ANN đã được nghiên cứu và ứng dụng rộng rãi trong hầu hết các ngành khoa học hiện nay. Trong dự báo chất lượng không khí, ANN cũng đã đạt được nhiều thành tựu đáng chú ý và đang trở thành một giải pháp hữu hiệu để kết hợp cũng như thay thế các mô hình dự báo số trị trong một số trường hợp như dự báo ô nhiễm, khí thải dữ liệu, chất lượng nguồn nhân lực thấp và cơ sở hạ tầng không đồng bộ. Tuy nhiên, tại Việt Nam trên cơ sở dữ liệu mờ, việc nghiên cứu vấn đề này còn khá ít, điều này cho thấy hướng tiếp cận này còn chưa nhận được sự quan tâm đúng mức. Vì vậy, bài báo này cung cấp một cái nhìn tổng quan về tiềm năng ứng dụng của ANN trong dự báo chất lượng không khí trên Thế giới và Việt Nam, các phân tích về thành tựu cũng như một số khó khăn khi ứng dụng công cụ này để xây dựng mô hình dự báo chất lượng không khí cũng được đề cập. Từ đó có thể góp phần vào việc thúc đẩy hơn nữa hướng nghiên cứu này tại Việt Nam.

**Từ khóa:** Dự báo, mạng nơron nhân tạo, ANN, chất lượng không khí, Việt Nam

## MỞ ĐẦU

Cũng giống như thời tiết, chất lượng không khí ảnh hưởng trực tiếp đến sức khỏe con người. Khi nồng độ các chất ô nhiễm vượt quá ngưỡng cho phép, tiếp xúc với thời gian ngắn có thể gây ra các phản ứng cấp tính như giảm tầm nhìn, khó thở, cay mắt..., tiếp xúc với thời gian dài có thể gây ra các bệnh慢 tính về hô hấp, tim mạch và có thể cả ung thư. Ngoài ra, ô nhiễm không khí còn gây ra các ảnh hưởng xấu đến các hệ sinh thái tự nhiên [1]. Do đó, bên cạnh công tác quan trắc thì dự báo được nồng độ các chất ô nhiễm không khí, cũng như diễn biến của chúng có ý nghĩa vô cùng quan trọng trong việc đánh giá các mối nguy cơ tiềm tàng của ô nhiễm không khí. Chính vì vậy, trong hơn một thập kỷ trở lại đây, các nghiên cứu ứng dụng các mô hình dự báo chất lượng không khí tại Việt Nam đã tăng lên đáng kể và đạt được nhiều thành tựu.

Hiện nay, các mô hình dự báo chất lượng không khí là rất đa dạng, phụ thuộc vào phương pháp tiếp cận, công cụ sử dụng... Tuy nhiên, trừ chung lại chúng có thể chia

thành hai nhóm là mô hình số trị (numerical model) và mô hình thống kê (statistical model). Mô hình số trị có ưu điểm là độ chính xác cao và dự báo được trên phạm vi rộng (quốc gia, vùng, miền...), tuy nhiên, chúng đòi hỏi một lượng lớn thông tin về nguồn gây ô nhiễm, các quá trình vật lý và hóa học có ảnh hưởng đến diễn biến của chất ô nhiễm cần dự báo, điều mà trong thực tế không thể đáp ứng được. Do đó, trong thực tế các mô hình này đều sử dụng các phương pháp giới hạn điều kiện biến.Thêm vào đó, do tính phức tạp và khối lượng tính toán lớn nên mô hình số trị còn đòi hỏi cao về cơ sở hạ tầng. Trong khi đó, mô hình thống kê mặc dù không mạnh trong dự báo trên phạm vi rộng và độ chính xác không cao trong trường hợp nồng độ của chất ô nhiễm cần dự báo có biến động lớn, nhưng việc xây dựng và vận hành chúng khá đơn giản, vì vậy, không đòi hỏi cao về trình độ chuyên môn của người sử dụng cũng như yêu cầu về cơ sở hạ tầng là đơn giản hơn. Tuy nhiên, hiệu suất của mô hình thống kê phụ thuộc vào số lượng cũng như chất lượng của dữ liệu quan trắc chất lượng không khí, cho nên, khi áp dụng vào thực tế cần phải xem xét kỹ lưỡng vấn đề này [1].

\* Tel: 0912 901524, Email: macdh@mtut.edu.vn

Tại Việt Nam, các mô hình đã được nghiên cứu và ứng dụng mới chỉ tập trung vào nhóm mô hình số trị. Trong khi đó trên thế giới, bên cạnh các mô hình số trị, các mô hình dự báo thống kê cũng đã được ứng dụng rộng rãi thành công tại nhiều quốc gia [2], [4], nó giải quyết được một số hạn chế của mô hình số trị. Trong số các công cụ được ứng dụng để xây dựng mô hình dự báo thống kê chất lượng không khí như RM (Regression method), FL (Fuzzy logic), ARIMA... thì mạng nơron nhân tạo (ANN – artificial neural network) được ứng dụng rộng rãi hơn cả, và đã đạt được nhiều thành tựu đáng chú ý [3], [5], [6]. Tuy nhiên, cho đến nay các nghiên cứu về vấn đề này tại Việt Nam vẫn còn rất hạn chế, điều này có thể thấy rõ trên thư viện học liệu mở. Chính vì vậy, bài báo này tập trung vào giới thiệu về những thành tựu đã đạt được của công cụ ANN trong lĩnh vực dự báo chất lượng không khí nhằm thúc đẩy hơn nữa hướng nghiên cứu này tại Việt Nam.

## MẠNG NƠRON NHÂN TẠO VÀ ỨNG DỤNG TRONG DỰ BÁO CHẤT LƯỢNG KHÔNG KHÍ

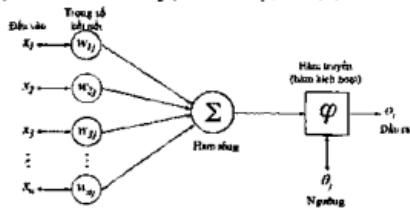
### Giới thiệu về mạng nơron nhân tạo (Artificial Neural Networks – ANNs)

Cho đến nay, có rất nhiều phương pháp tiếp cận để xây dựng các mô hình dự báo thống kê, trong đó có dự báo chất lượng không khí đã được bao cáo trong rất nhiều tài liệu chuyên ngành. Tuy nhiên, việc lựa chọn phương pháp tiếp cận phù hợp nhất phụ thuộc vào độ phức tạp của vấn đề cần dự báo. Giả sử trong trường hợp dữ liệu đầy đủ (chất lượng tốt), tài nguyên về hệ thống máy tính dù mạnh, và sự hiểu biết lý thuyết về các quá trình là dù sâu sắc, thì một mô hình dự báo đang chính xác có lẽ là một trong những sự lựa chọn tốt nhất. Tuy nhiên, cũng giống như việc dự báo kinh tế hay dự báo thời tiết thì các tham số có ảnh hưởng đến một thông số chất lượng không khí (chất ô nhiễm) là vô cùng phức tạp, nó bao gồm tổng hợp các yếu tố khí tượng (tốc độ gió, hướng gió, nhiệt độ,

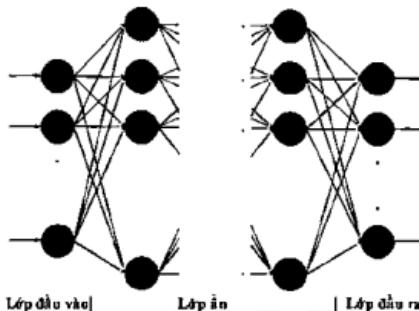
độ ẩm, mưa...) và sự biến đổi hóa học của các tiền chất ô nhiễm, cũng như của chính nó trong không khí. Chính vì vậy, việc xây dựng một mô hình chính xác có thể biểu diễn được đầy đủ các mối quan hệ này có thể nói là bất khả thi. Từ cuối thập niên 90 của thế kỷ XX đến nay, các ứng dụng của mạng nơron nhân tạo (ANN) trong dự báo chất lượng không khí đã được chứng minh là giải pháp thay thế hữu hiệu các mô hình thống kê truyền thống [1], bởi ANNs có thể được huấn luyện để mô tả gần như hoàn hảo một hàm số bất kỳ. Khác với các mô hình thống kê truyền thống, ANNs không đòi hỏi sự hiểu biết lý thuyết về quá trình cũng như phân bố dữ liệu, thông qua quá trình huấn luyện ANNs có thể từ khái quát chính xác các thông tin về các mối quan hệ giữa các biến số ẩn trong bộ dữ liệu huấn luyện ngay cả trong những vấn đề mới và chưa xuất hiện các mô hình lý thuyết. Chính vì những khả năng này, ANNs đã trở thành một đề tài hấp dẫn để phát triển các mô hình số và cân nhắc khi lựa chọn các phương pháp thống kê. Và trong lĩnh vực dự báo chất lượng không khí, như sẽ thấy ở phần sau, việc ứng dụng ANNs đã đạt được nhiều thành tựu đáng chú ý cả trên Thế giới và Việt Nam.

Mạng nơron nhân tạo (ANN) là một mô hình toán được xây dựng dựa trên việc mô hình hóa quá trình xử lý thông tin của một hệ nơron sinh học, bao gồm ba hoặc nhiều lớp chứa các nơron (tế bào nơron) được mô tả trong Hình 2. Trong đó, lớp đầu tiên được gọi là lớp đầu vào (input layer), lớp cuối cùng chứa kết quả được gọi là lớp đầu ra (output layer) và các lớp còn lại ở giữa được gọi là lớp ẩn (hidden layers). Mỗi nơron nhận tín hiệu từ các nơron khác, mức độ quan trọng của các liên kết này được thể hiện dưới dạng một số thực – gọi là trọng số liên kết. Tập hợp các trọng số này tạo thành ma trận trọng số  $w_{ij}$ , nó mã hóa kiến thức, kinh nghiệm tích lũy được của ANN thông qua quá trình học tập từ tập dữ liệu đầu vào. Ưu điểm lớn nhất của ANN chính là khả năng tự phát hiện các thông tin về mối quan hệ giữa các thông số

đầu vào và đầu ra là trong bộ dữ liệu, mà chúng không thể mô tả được bằng các phương trình toán học chính xác, về cơ bản, quá trình này được thực hiện dựa trên việc điều chỉnh trọng số liên kết giữa các nơron – được gọi là quá trình huấn luyện của một ANN.



Hình 1. Mô hình toán của một nơron



Hình 2. Kiến trúc cơ bản của một ANN

Quá trình xử lý thông tin của nơron thứ  $j$  (Hình 1) được mô tả vắn tắt theo phương trình (1) và (2). Đầu tiên, nó sẽ nhận tín hiệu  $x_i$  từ các nơron phía trước có trọng số tương ứng là  $w_{ij}$ . Toàn bộ các thông tin này sẽ được tổng hợp thông qua hàm tổng (summing function) sau đó được xử lý bởi hàm truyền (transfer function) để chuyển thành tín hiệu đầu ra.

$$\varphi(x) = \sum_{i=1}^n w_{ij} x_i + \theta_j \quad (1)$$

$$output = f\left(\sum_{i=1}^n w_{ij} x_i + \theta_j\right) \quad (2)$$

Ứng dụng của ANN trong dự báo chất lượng không khí

Về mặt nguyên tắc, ANN dự đoán hành vi của đối tượng (kết quả) trong tương lai dựa vào chức năng suy luận của nó thông qua việc phân tích và khai quát thông tin từ bộ dữ liệu

kinh nghiệm trong quá khứ và hiện tại - giống với cơ chế của quá trình dự báo. Như đã biết, mối quan hệ giữa các tiền chất ô nhiễm và các thông số khí tượng đến chất ô nhiễm cần dự báo là rất phức tạp, do đó, có thể nói dự báo chất lượng không khí là vô cùng khó khăn.

### Thế giới

Ngay từ thập niên 90 của thế kỷ XX, một số mô hình ứng dụng ANN trong dự báo nồng độ chất ô nhiễm đã thu được một số thành công bước đầu gây tiếng vang lớn. Có thể kể đến ở đây như [3], trong nghiên cứu này Yi và Prybutok đã phát triển thành công mô hình dự báo nồng độ ôzôn cho khu công nghiệp tại Bắc Mỹ. Mô hình sử dụng chín thông số đầu vào bao gồm nồng độ  $O_3$ , tinh chất của ngày cần dự báo (ngày nghỉ/làm việc), nồng độ của bốn tiền chất ( $CO_2$ ,  $NO$ ,  $NO_2$  và  $NO_x$ ) và bốn thông số khí tượng (nhiệt độ, tốc độ gió, hướng gió). Kết quả thu được từ mô hình ANN so với hai mô hình đối chứng một sử dụng phương pháp phân tích hồi quy phi tuyến (regression analysis) và một là mô hình ARIMA đều xuất sắc hơn. Dựa trên những thành công đó, cho đến nay ANN đã được nghiên cứu và ứng dụng rộng rãi tại nhiều quốc gia trên Thế giới để dự báo nhiều thông số ô nhiễm khác nhau và đã đạt được những thành tựu nhất định về mặt kỹ thuật cũng như xã hội.

Năm 2003, Chaloulakou đã so sánh hiệu năng của mô hình ANN (MLP) và hồi quy đa biến (LR - Multiple Linear Regression) khi phát triển mô hình dự báo thông kê nồng độ bụi PM<sub>10</sub> cho thành phố Athen, Hy Lạp [6]. Kết quả thu được cho thấy, trong tất cả các trường hợp được so sánh với cùng thông số đầu vào, mô hình ANN đều cho kết quả xuất sắc hơn với tất cả các chỉ tiêu thống kê được xem xét. Cụ thể, với năm thông số đầu vào ( $DT = (T_{max} - T_{min})$ ,  $WS$  – tốc độ gió,  $WDI$  – hướng gió,  $RH$  – độ ẩm tương đối, và  $DoW$  là tham số ngày trong tuần) hệ số MAE, RMSE,  $R^2$  và  $d$  của mô hình MLP1 lần lượt là  $16(1,48)$   $\mu g/m^3$ ,  $21,19(0,95)$   $\mu g/m^3$ ,  $0,47(0,03)$  và

0,78(0,01), trong khi đó của LR1 tương ứng lần lượt là 17,96(1,6)  $\mu\text{g}/\text{m}^3$ ; 23,4(1,12)  $\mu\text{g}/\text{m}^3$ , 0,34(0,03) và 0,73(0,01). Khi bổ sung thêm nồng độ của  $\text{PM}_{10}$  tại bước thời gian trước đó ( $\text{PM}_{10-1}$ ) hiệu suất dự báo của cả mô hình LR và MLP đều có sự cải thiện và cũng giống như MLP1 thì mô hình MLP2 cũng cho hiệu suất tốt hơn trên bộ dữ liệu kiểm chứng (testing), kết quả MAE, RMSE,  $R^2$  và d lần lượt là 12,62(1,21)  $\mu\text{g}/\text{m}^3$ , 16,94(0,76)  $\mu\text{g}/\text{m}^3$ , 0,65(0,03) và 0,9(0,01). Trong khi LR2 là 14,07(1,26)  $\mu\text{g}/\text{m}^3$ , 18,37(0,88)  $\mu\text{g}/\text{m}^3$ , 0,6(0,03) và 0,87(0,01). Kết quả này tương đồng với các nghiên cứu tương tự đã được báo cáo trên thế giới, tại Bi [5], Italy [7], Trung Quốc [8], Chile [9]... Thêm vào đó, các kết quả được báo cáo trong [10] khi so sánh hiệu năng của một số mô hình dự báo thống kê đối với nồng độ  $\text{SO}_2$  tại hai thành phố Siracura (Italy – thành phố công nghiệp) và Belfast (Anh – khu đô thị), kết quả thu được từ mô hình ANN với các mô hình khác là rất khả quan. Cụ thể, đối với Siracura (dữ liệu trạm Melilli) hệ số xác định ( $d$ ) và hệ số tương quan Pearson ( $r$ ) lần lượt là 0,77 và 0,66 tương ứng với các mô hình khác đạt 0,67 đến 0,78 đối với hệ số  $d$  và 0,53 đến 0,64 đối với  $r$ . Khi dự báo cho Belfast kết quả của mô hình ANN tương ứng là 0,83 đối với  $d$  và 0,75 đối với  $r$ .

Đối với trường hợp dự báo cùng lúc nhiều thông số ô nhiễm [11], kết quả được báo cáo trong nhiều nghiên cứu đã cho thấy điểm mạnh của công cụ ANN so với các phương pháp cổ điển như hồi quy (MLR), logic mờ (FR),... do việc xây dựng mô hình khá đơn giản nhưng hiệu suất lại tương đối cao, không đòi hỏi cao về nguồn nhân lực cũng như cơ sở hạ tầng. Đặc biệt, đối với các thông số ô nhiễm thứ cấp như  $\text{O}_3$  thì các mô hình MLR, FL, ARIMA... không đạt được hiệu suất cao do nồng độ của chất ô nhiễm thứ cấp  $\text{O}_3$  ngoài phụ thuộc vào các tham số khí tượng, nó còn phụ thuộc vào diễn biến của các tiền chất (đối với  $\text{O}_3$  là  $\text{NO}_x$ ,  $\text{CO}$  và VOC). Năm 2007 U.Brunelli và cộng sự [7] đã phát triển mô

hình ANN để dự báo năm thông số chất lượng không khí (AQI – air quality index) cho thành phố Palermo, Italy, bao gồm  $\text{SO}_2$ ,  $\text{O}_3$ ,  $\text{PM}_{10}$ ,  $\text{NO}_2$  và  $\text{CO}$  sử dụng dữ liệu từ 08 trạm quan trắc. Mô hình ELMAN – ANN đã được lựa chọn với MAE của  $\text{SO}_2$  nằm trong khoảng 1,52  $\mu\text{g}/\text{m}^3$  – 3,52  $\mu\text{g}/\text{m}^3$  tương ứng với  $r$  nằm trong khoảng 0,89 – 0,96;  $\text{CO}$  là 0,09  $\mu\text{g}/\text{m}^3$  – 0,39  $\mu\text{g}/\text{m}^3$  và 0,93 – 0,96;  $\text{NO}_2$  là 3,77  $\mu\text{g}/\text{m}^3$  – 23,42  $\mu\text{g}/\text{m}^3$  và 0,90 – 0,98;  $\text{O}_3$  là 3,29  $\mu\text{g}/\text{m}^3$  – 3,59  $\mu\text{g}/\text{m}^3$  và 0,72 – 0,87; và  $\text{PM}_{10}$  lần lượt là 2,77  $\mu\text{g}/\text{m}^3$  – 5,58  $\mu\text{g}/\text{m}^3$  và 0,93 – 0,97 tương ứng. Kết quả từ nghiên cứu này một lần nữa khẳng định việc dự báo các chất ô nhiễm thứ cấp như  $\text{O}_3$  là khó khăn hơn, chính vì vậy độ tin cậy của mô hình đối với thông số này không cao như các thông số ô nhiễm sơ cấp khác. Chính vì vậy, nhiều mô hình lai đã được nghiên cứu nhằm nâng cao hiệu suất dự báo các chất ô nhiễm thứ cấp, đặc biệt là  $\text{O}_3$ , có thể kể đến ở đây như mô hình kết hợp giữa PCR (hồi quy thành phần chính) và ANN được báo cáo trong [12], mô hình kết hợp giữa GA (thuật giải di truyền) và ANN [13], PCA – ANN [2], [14]. Kết quả được báo cáo trong các tài liệu này cho thấy hiệu suất dự báo được cải thiện đáng kể so với ANN thông thường.

Với điều kiện khí hậu tương tự Việt Nam, các mô hình dự báo chất lượng không khí ứng dụng ANN cũng đã được nghiên cứu và ứng dụng tại Thái Lan [15] đã phát triển mô hình ANN và LR để dự báo nồng độ  $\text{O}_3$  cho Bang Cốc, Thái Lan. Báo cáo này cho thấy hiệu năng của mô hình ANN là xuất sắc hơn rất nhiều so với LR với cùng tham số đầu vào, cụ thể các chỉ số  $R^2$  của mô hình ANN đạt 0,85 – 0,89, tương ứng với RMSE = 11,8 ppb – 13,5 ppb và MAE = 8,6 ppb – 10,3 ppb. Trong khi đó, chỉ tiêu  $R^2$  của mô hình LR cao nhất chỉ đạt 0,396. Các kết quả này cho thấy, ANN là một hướng tiếp cận đơn giản và khả thi.

#### *Việt Nam*

Tại Việt Nam, mặc dù vẫn để dự báo chất lượng không khí cũng đã được nghiên cứu từ thập niên 90 của thế kỷ XX, tuy nhiên, các

mô hình đã được nghiên cứu và sử dụng tập trung chủ yếu vào nhóm các mô hình dự báo số trị như CMAQ, WRF/Chem... [16], [18]. Đến năm 2015, vẫn đề nghiên cứu ứng dụng ANN trong dự báo chất lượng không khí mới có những báo cáo đầu tiên về vấn đề này được xuất bản [19]. Trong nghiên cứu này, tác giả nghiên cứu sử dụng ANN để bù đũ liệu quan trắc chất lượng không khí của bụi PM<sub>10</sub> và SO<sub>2</sub>, kết quả thu được bước đầu là khá khá quan với hệ số tương quan Pearson ( $r$ ) nằm trong khoảng 0,39 đến 0,51. Thêm nữa, quy luật phân bố xác suất của bộ dữ liệu đã bù với bộ dữ liệu thực tế có sự tương quan tốt và tương đồng với các nghiên cứu về phân bố xác suất trước đó [20]. Từ đó cho đến nay, một số dạng ANNs cũng đã được nghiên cứu nhằm nâng cao hiệu suất dự báo và đảm bảo tính khả thi khi ứng dụng các mô hình này trong thực tế. Mặc dù số lượng các nghiên cứu còn ít, nhưng kết quả thu được cũng đạt được một số thành tựu đáng chú ý. [21] đã nghiên cứu ứng dụng mạng ESN (Echo state network – mạng trạng thái phản hồi) và MLP để dự báo nồng độ của ba thông số chất lượng không khí, bao gồm SO<sub>2</sub>, PM<sub>10</sub> và NO<sub>2</sub> tại Hà Nội với 07 biến đầu vào bao gồm nồng độ của ba chất ô nhiễm cần dự báo (SO<sub>2</sub>, PM<sub>10</sub> và NO<sub>2</sub>) và bốn thông số khí tượng gồm tốc độ gió (WS), hướng gió (WD), nhiệt độ (T) và độ ẩm (RH) tại bước thời gian  $t-1$ . Độ chính xác của mô hình ESN đối với SO<sub>3</sub> trung bình đạt 74,7% với bước dự báo 07 ngày, với PM<sub>10</sub> đạt 83,8% và NO<sub>2</sub> là 60%. Còn MLP tương ứng là 76,5%, 77,6% và 72,7%, tương ứng.

Các kết quả dự báo đa thông số cũng đã được báo cáo trong [22]. Trong nghiên cứu này, tác giả sử dụng MLP để xây dựng mô hình dự báo nồng độ ôzôn với bước thời gian 08 giờ. Cấu trúc của MLP được tối ưu hóa bằng thuật toán di truyền (GA). Chín biến đầu vào được sử dụng bao gồm O<sub>3</sub>, NO, NO<sub>2</sub>, VOC, tốc độ gió (WS), hướng gió (WD), nhiệt độ (T), cường độ bức xạ cực tím (UV) và thời gian ( $t$ ). Mô hình mà tác giả sử dụng ngoài việc dự báo thông số chính là O<sub>3</sub> thì các tham số về các tiền chất ô nhiễm gồm (NO, NO<sub>2</sub> và

VOC) cũng được dự báo đồng thời và được sử dụng như là đầu vào cho bước dự báo kế tiếp. Với bước dự báo 08 giờ, hệ số xác định R<sup>2</sup> của mô hình đạt 0,54 tương ứng với RMSE và MAE lần lượt là 5,37 ppb và 4,49 ppb. Trong khi mô hình LR được sử dụng để so sánh là 0,54 với RMSE và MAE là 22,83 ppb và 20,96 ppb. Đối với các thông số phụ gồm VOC, NO và NO<sub>2</sub> hệ số R<sup>2</sup> đạt tương ứng là 0,81; 0,69 và 0,83. Kết quả của các nghiên cứu này là khá tương đồng với các kết quả của các mô hình cùng loại đã được báo cáo trên Thế giới, kèm theo đó, diễn biến nồng độ của các thông số được dự báo có sự tương quan tốt với diễn biến thực do tai trạm quan trắc. Điều này là một tin hiệu tốt, minh chứng rằng ANN là một công cụ có tính khả thi cao với điều kiện dữ liệu cũng như nguồn nhân lực quản lý chất lượng không khí hiện nay tại Việt Nam.

## KẾT LUẬN VÀ KIẾN NGHI

Trong dự báo chất lượng không khí, diễn biến nồng độ của các thông số cần dự báo là một mối quan hệ phức tạp giữa các quá trình vật lý (khuếch tán, lắng, hòa tan...) và các quá trình biến đổi hóa học, một mô hình dự báo tốt là một mô hình có thể biểu diễn được tất cả các mối quan hệ này. Tuy nhiên trên thực tế, một mô hình lý thuyết khó có thể thực hiện được. Trong khi, ANN từ lâu đã chứng minh được khả năng của mình trong lĩnh vực này bởi ANN không đòi hỏi phải xây dựng mô hình lý thuyết mà bản thân nó có khả năng phát hiện và ánh xa các mối quan hệ ẩn trong dữ liệu thông qua quá trình học. Các kết quả thu được từ các nghiên cứu về ứng dụng của ANN trong dự báo chất lượng không khí trên đã cho thấy, ANN là một hướng tiếp cận khả thi. Đặc biệt trong điều kiện Việt Nam, việc triển khai các mô hình số trị trên diện rộng là khá khó khăn do số lượng cũng như chất lượng nguồn nhân lực, mặt khác, loại mô hình này đòi hỏi cao về dữ liệu cũng như cơ sở hạ tầng. Trong khi, việc vận hành các mô hình dự báo thông kê, đặc biệt là ANN khá đơn giản, yêu cầu về dữ liệu và cơ sở hạ tầng cũng

không quá cao. Tuy nhiên cũng cần phải nói thêm, hiệu năng của các mô hình ANN phụ thuộc vào số lượng và chất lượng của dữ liệu quan trắc, chính vì vậy, việc triển khai rộng rãi loại mô hình này thực tế cũng sẽ gặp các khó khăn nhất định bởi hệ thống quan trắc tại Việt Nam còn ít và thiếu đồng bộ. Tuy vậy, sự hiện diện của các mô hình dự báo thông kê nói chung và ANN nói riêng sẽ đáp ứng và thay thế được mô hình dự báo số trị trong các trường hợp dữ liệu không đầy đủ và dự báo điểm ô nhiễm (dự báo phạm vi hẹp).

Kết quả nghiên cứu ứng dụng ANN trong dự báo chất lượng không khí từ năm 1996 đến nay trên Thế giới và Việt Nam được xem xét trong bài báo này phần nào mô tả được bức tranh toàn cảnh về khả năng ANN trong lĩnh vực này. Kết quả của các nghiên cứu này cho thấy, ANN là một công cụ mạnh và hoàn toàn khả thi để ứng dụng tại Việt Nam. Tuy nhiên, các nghiên cứu tại Việt Nam về hướng tiếp cận này mặc dù cũng đã đạt được một số kết quả nhất định nhưng còn khá hạn chế, vì vậy mong muốn với những thông tin cung cấp từ bài báo này sẽ đầy mạnh hơn nữa việc nghiên cứu hướng tiếp cận này trong dự báo chất lượng không khí nói riêng và bảo vệ môi trường nói chung.

#### TÀI LIỆU THAM KHẢO

- Yang Zhang, Marc Bocquet, Vivien Mallet, Christian Seigneur, and Alexander Baklanov (2012), "Real-time air quality forecasting, part I: History, techniques, and current status", *Atmospheric Environment* 60, pp 632-655
- A. S Luna, M. L L Paredes, G. C. G de Oliveira, and S M Correa (2014), "Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil", *Atmospheric Environment* 98, pp 98-104
- Junsuk Yi, Victor R Prybutok (1996), "A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area", *Environmental Pollution* 92 (3), pp 349-357.
- Alessandro Fassò and Ilia Negri (2002), "Multi-step forecasting for nonlinear models of high frequency ground ozone data: a Monte Carlo approach", *Environmetrics* 13 (4), pp 365-378
- Hooyberghs Jef, Mensink Clemens, Dumont Gerwin, Fierens Frans, and Brasseur Olivier (2005), "A neural network forecast for daily average PM10 concentrations in Belgium", *Atmospheric Environment* 39 (18), pp 3279-3289
- Archontoulou Chaloulakou, Georgios Grivas and Nikolas Spyrellis (2003), "Neural Network and Multiple Regression Models for PM10 Prediction in Athens: A Comparative Assessment", *Journal of the Air & Waste Management Association* 53, pp 1183-1190.
- U. Brunelli, V. Piazza, L. Pignato, F. Sorbello, and S. Vitabile (2007). "Two-days ahead prediction of daily maximum concentrations of SO<sub>2</sub>, O<sub>3</sub>, PM10, NO<sub>2</sub>, CO in the urban area of Palermo, Italy", *Atmospheric Environment* 41 (14), pp. 2967-2995
- Wei-Zhen Lu, Wen-Jian Wang (2005), "Potential assessment of the "support vector machine" method in forecasting ambient air pollutant trends", *Chemosphere* 59, pp. 693-701
- Patricio Perez, Jorge Reyes (2006), "An integrated neural network model for PM10 forecasting", *Atmospheric Environment* 40, pp 2845-2851.
- Giuseppe Nunnari, et al. (2004), "Modelling SO<sub>2</sub> concentration at a point with statistical approaches", *Environmental Modelling & Software* 19, pp 887-905.
- Kunwar P Singh, Shikha Gupta, Atulesh Kumar, and Sheo Prasad Shukla (2012), "Linear and nonlinear modeling approaches for urban air quality prediction", *Science of the Total Environment* 426, pp. 244-255
- Saleh M. Al-Alawi, Sabah A. Abdul-Wahab, and Charles S. Bakheit (2008), "Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone", *Environmental Modelling & Software* 23, pp 396-403.
- Yu Feng, Wenfang Zhang, Dezhong Sun, and Liqiu Zhang (2011), "Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification", *Atmospheric Environment* 45 (11), pp. 1979-1985
- Che-hui Tsai, Li-chiu Chang, and Hsu-cherng Chiang (2009), "Forecasting of ozone episode days by cost-sensitive neural network methods", *Science of The Total Environment* 407 pp 2124-2135
- Le Hoang Nghiêm and Nguyễn Thị Kim Oanh (2009), "Comparative analysis of maximum

daily ozone levels in urban areas predicted by different statistical models". *Science Asia* 35, pp 276-283.

16 Trương Anh Sơn and Dương Hồng Sơn (2007), "Nghiên cứu thử nghiệm áp dụng hệ thống mô hình dự báo chất lượng không khí da quy mô CMAQ ở Việt Nam", *Tạp chí Khoa học Tự nhiên*, tháng 12/2007, pp. 43-49

17 Dam Duy An, Hoang Xuan Co, and Nguyen Thi Kim Oanh (2008), "Photochemical smog introduction and episode selection for the ground-level ozone in Hanoi, Vietnam", *VNU Journal of Science, Earth Sciences* 24, pp 169-175

18 Le Hoang Nghiêm and Nguyen Thi Kim Oanh (2009), "Regional-scale modeling ozone air quality over the continental south east asia", *Tạp chí phát triển KH&CN*, 12 (2), pp. 111-120

19 Mac Duy Hung, Nghiêm Trung Dũng, and Dinh Thu Hang (2015), "Application of artificial

neural network to fill in the missing monitoring data of air quality", *Journal of Science and Technology (VAST)* 53 (3A), pp 199-204.

20 Chủ Thị Hồng Nhụng, Nghiêm Trung Dũng (2012), "Xác định luật phân bố xác suất của dữ liệu chất lượng không khí được quan trắc tại Hà Nội", *Tạp chí Khoa học và Công nghệ*, 50 (1), pp 81-87.

21 Mac Duy Hung and Nghiêm Trung Dũng (2016), "Application of Echo State Network for the forecast of air quality", *Journal of Science and Technology (VAST)* 54 (1), pp 54-63

22. Mac Duy Hung, Nghiêm Trung Dũng, and Hoang Xuan Co (2016), "Application of Multilayer Perceptron Neural Network for the forecast of tropospheric ozone in Hanoi", *Journal of Science and Technology of Six University* 111, pp 46-51

## SUMMARY

### A REVIEW OF APPLICATIONS OF ARTIFICIAL NEURAL NETWORKS TO FORECAST OF AIR QUALITY

Nghiem Trung Dung<sup>1</sup>, Mac Duy Hung<sup>1,2\*</sup>, Hoang Xuan Co<sup>3</sup>

<sup>1</sup>School of Environmental Science and Technology – HUST

<sup>2</sup>College of Technology – TNU

<sup>3</sup>Hanoi University of Sciences – VNU

Artificial neural networks (ANN) has become an effective approach to alternate the traditional statistical model. Thus, ANNs have been successfully applied in many scientific disciplines. In the field of forecast of air quality, ANNs have had many notable achievements, therefore, ANNs have become an effective solution to combine and alternate the numerical forecasting models in some cases such as the forecast of air quality of pollution point, lack of data, low quality of human resources and asynchronous infrastructure. Nevertheless, in the best of our knowledge only small number of studies conducted in Vietnam on the use of ANN for forecast of air quality. These indicated that, this approach is still new and has not been received interest in Vietnam. This study is, therefore, aimed at the providing an overview of the potential applications of ANN in forecasting air quality in the world and Vietnam. Thence, it can contribute to furthering research of this field in Vietnam.

**Keywords:** Artificial neural network, ANN, forecast, prediction, air quality.