

*Bài báo khoa học*

## **Nghiên cứu công nghệ dự báo mưa AI thí điểm tại TP. Hồ Chí Minh**

**Phạm Thanh Long<sup>1</sup>, Lê Văn Phận<sup>2</sup>, Nguyễn Phương Đông<sup>1</sup>, Lê Hồng Dương<sup>3</sup>, Trần Tuấn Hoàng<sup>1</sup>**

<sup>1</sup> Phân viện KH Khí tượng Thủy văn và Biến đổi khí hậu; longpham.sihymete@gmail.com; hoangkttv@gmail.com; nguyenphuongdongkttv@gmail.com

<sup>2</sup> ĐH Nông Lâm TPHCM; lvphan@hcmuaf.edu.vn

<sup>3</sup> Cục Bảo vệ Môi trường miền Nam; hongduong.tcmt@gmail.com

\*Tác giả liên hệ: longpham.sihymete@gmail.com; Tel: +84-905779777

**Tóm tắt:** Các nước tiên tiến sử dụng AI trong dự báo thời tiết rất phổ biến và ngày càng chính xác hơn. Trong những năm gần đây hệ thống dữ liệu quan trắc cập nhật theo thời gian thực của các đài khí tượng thủy văn đã phát triển rất mạnh mẽ, nguồn dữ liệu lớn và liên tục, đây là thuận lợi rất lớn khi triển khai nghiên cứu về các AI trong dự báo mưa điểm thử nghiệm tại TP. Hồ Chí Minh. Các yếu tố nhiệt độ không khí, độ ẩm không khí, tốc độ gió, khí áp tại điểm quan trắc là dữ liệu đầu vào cho hệ thống dự báo mưa AI. Dựa trên cơ sở các trạm quan trắc thời tiết theo thời gian thực, công nghệ xử lý phần mềm AI trên máy tính rất nhanh, quá trình học dữ liệu và dự báo kết quả mới rất nhanh và thuận lợi. Nghiên cứu đã tiến hành thử nghiệm và dự báo với độ chính xác khá cao. Kết quả dự báo thời điểm trước khi có mưa và kết quả dự báo thời điểm có mưa tại 03 trạm thuduc, quan2, quan9 cho thấy kết quả dự báo chính xác cao nhất với hệ số tương quan lần lượt là 0,93, 0,78 và 0,7.

**Từ khóa:** Dự báo mưa AI; Dữ liệu lớn; Khoa học dữ liệu; Trí tuệ nhân tạo.

### **1. Mở đầu**

Trên thế giới Google đã triển khai hệ thống dự báo ngập lũ ứng dụng trí tuệ nhân tạo ở Banlades và Ấn độ năm 2018, đây là một hệ thống dữ liệu lớn và hoạt động rất hiệu quả, hệ thống triển khai trên khu vực dân cư gần 200 triệu người và có đến 65% số người nhận được thông tin cảnh báo ngập lũ từ năm 2018 đến năm 2020 trong vòng 48 giờ trước khi xảy ra nguy hiểm [1].

Trong nước hệ thống cảnh báo lũ trực tuyến Sông Vu Gia–Thu Bồn [2] đã triển khai nghiên cứu trực tuyến cập nhật dữ liệu dự báo và chạy mô hình ngập lũ phạm vi rộng lớn tại các tỉnh Quảng Nam và TP. Đà Nẵng, trong nghiên cứu này đã sử dụng dữ liệu thời gian thực nhưng chưa ứng dụng công nghệ AI hỗ trợ cảnh báo nhanh các vùng ngập [2]. Dự báo mưa dùng công nghệ trí tuệ nhân tạo AI tại Việt Nam do nhóm dự án nghiên cứu cảnh báo ngập sử dụng công nghệ AI tại TP.HCM năm 2020–2021 là một bước tiến mới trong tiến bộ khoa học công nghệ ngành Khí tượng thủy văn nước nhà.

Nhu cầu dự báo mưa dùng AI học dữ liệu quan trắc thời gian thực tại điểm nghiên cứu và dự báo thời đoạn ngắn (trước 3h) với độ chính xác cao. Đây là một ứng dụng đi trước một bước khi ngành khí tượng thủy văn chuyển dần các quan trắc thông thường sang quan trắc tự động, dữ liệu theo thời gian thực, các trạm quan trắc ngày càng nhiều, từ đó khả năng ứng dụng phần mềm này ngày càng lớn [3].

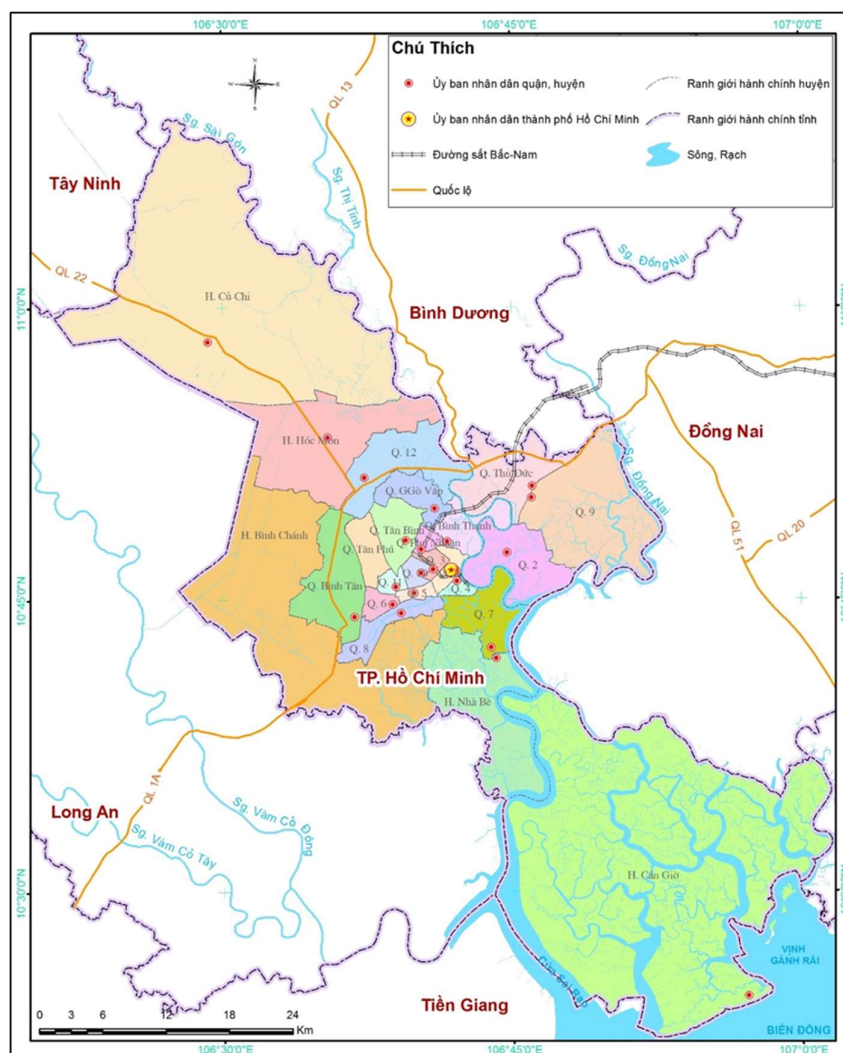
Công nghệ dự báo AI tự thu thập dữ liệu chạy dự báo và kết xuất ra các dữ liệu đầu vào cho các ứng dụng khác từ đó hệ thống có thể phát triển ứng dụng rộng rãi trong các lĩnh vực:

cảnh báo ngập, cung cấp tiêu thời tiết nhanh trong phạm vi hẹp, kịp thời cảnh báo các thời tiết cực đoan như gió lốc, mưa lớn, nhiệt cao, triều cường...[4]. Nghiên cứu đã thực hiện nghiên cứu thu thập dữ liệu và chạy thử các mô hình dự báo AI để chọn ra một mô hình tối ưu về độ chính xác và tốc độ xử lý phù hợp với điều kiện các máy móc thiết bị thời điểm hiện tại nghiên cứu [5].

## 2. Phương pháp nghiên cứu

### 2.1. Khu vực nghiên cứu

Thành phố Hồ Chí Minh nằm ở phía Tây Nam vùng Đông Nam Bộ. Phía Bắc giáp tỉnh Bình Dương, Tây Bắc giáp tỉnh Tây Ninh, Đông và Đông Bắc giáp tỉnh Đồng Nai, Đông Nam giáp tỉnh Bà Rịa–Vũng Tàu và biển Đông, Tây và Tây Nam giáp tỉnh Long An và Tiền Giang. TP.HCM nằm ở hạ lưu các con sông lớn: sông Đồng Nai, sông Sài Gòn, sông Bé, ven rìa Đồng bằng sông Cửu Long. TP.HCM nằm trong vùng có khí hậu nhiệt đới gió mùa mang tính chất cận xích đạo (Hình 1).



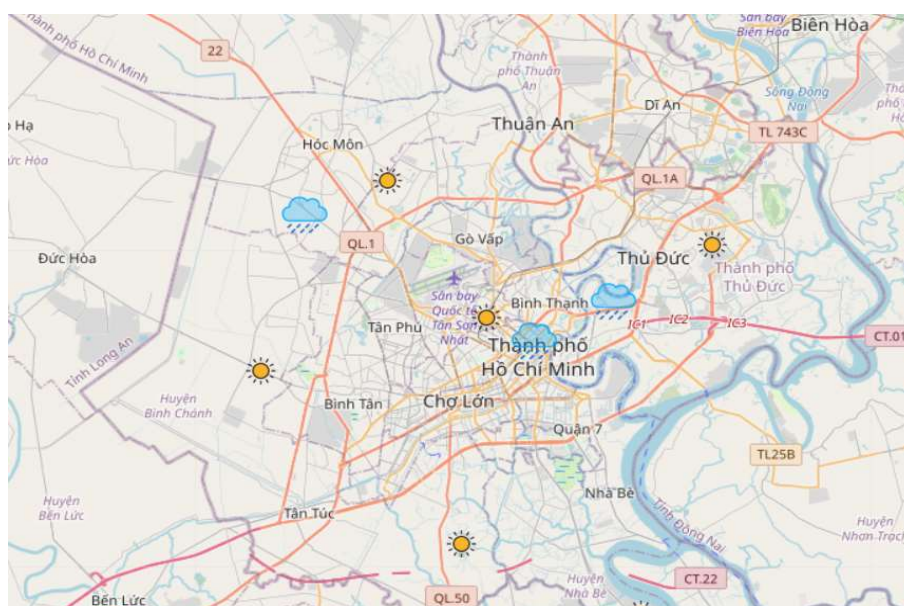
Hình 1. Bản đồ khu vực nghiên cứu.

Lượng bức xạ dồi dào, nắng trung bình 6,13 giờ/ngày. Nhiệt độ trung bình toàn năm khoảng 28,4°C. TP.HCM nằm trong khu vực khí hậu nhiệt đới gió mùa. Gió mùa mang một lượng lớn hơi ẩm từ phía Tây và Tây Nam. Với địa hình tự nhiên TP.HCM khá thấp cùng với

các mục đích sử dụng đất khác nhau đã dẫn đến sự phân bố rõ ràng về không gian và lượng mưa, thậm chí là phụ thuộc vào sự khác nhau của các quận nội thành của Thành phố. Tổng lượng mưa TP.HCM dao động từ 1.200–2.100 mm/năm [6–7]. Các trạm đo mưa tự động thuộc quản lý của Phân Viện Khoa học Khí tượng Thủy văn và Biến đổi khí hậu được sử dụng làm số liệu đầu vào cho nghiên cứu này chi tiết các trạm trình bày trong (Bảng 1, Hình 2).

**Bảng 1.** Danh sách trạm khí tượng thủy văn TP.HCM.

| Mã Trạm | Tên Trạm   | Địa danh                    | Kinh độ    | Vĩ độ     |
|---------|------------|-----------------------------|------------|-----------|
| 01      | 19 NTMK    | 19 NTMK                     | 106.699892 | 10.784652 |
| 02      | Quận 2     | Trần Ngọc Diện–Thảo điền    | 106.741254 | 10.806362 |
| 03      | Quận 9     | Đường 182–P.Tăng Nhơn phú A | 106.792684 | 10.836989 |
| 04      | Bình Chánh | Tỉnh Lộ 10 Bình Chánh       | 106.557478 | 10.768129 |
| 05      | Nhà Bè     | Đường số 3–KCN Hiệp Phước   | 106.755360 | 10.634954 |
| 06      | Hóc Môn    | Xuân Thới thượng 7–Hóc Môn  | 106.580568 | 10.853719 |
| 07      | Quận 12    | Tân Chánh Hiệp–Q.12         | 106.623837 | 10.871452 |
| 08      | Quận 8     | Mễ Cốc–Quận 8               | 106.632333 | 10.727641 |



**Hình 2.** Bản đồ vị trí các trạm Khí tượng Thủy văn TP.HCM.

## 2.2. Thuật toán vận dụng

*Random forest (RF) Algorithm* là thuật toán được các nhà khoa học công bố năm 1999 (L. Breiman) và được sử dụng rộng rãi trong trí tuệ nhân tạo, huấn luyện máy học có hiệu suất cao. Thuật toán được xây dựng trong thư viện scikit-learn của ngôn ngữ lập trình Python [8–10].

Random forest có hai dạng dự báo một là classifier theo dạng phân loại rẽ nhánh, hai là dạng Regression: thuật toán theo dạng hồi qui nơ tron [12–13].

Các bộ dữ liệu học: (*Training set*) là tập hợp mảng số liệu theo bộ:

Tree 1 = [A,B,C,D], Tree 2 = [A,B,D,C]...., và mảng kết quả  $R = [R_1, R_2, \dots]$

Thuật toán cơ bản như sau:

$$I(p, n) = \frac{-p}{p+n} \log_2 \left( \frac{p}{p+n} \right) - \frac{n}{p+n} \log_2 \left( \frac{n}{p+n} \right) \quad (1)$$

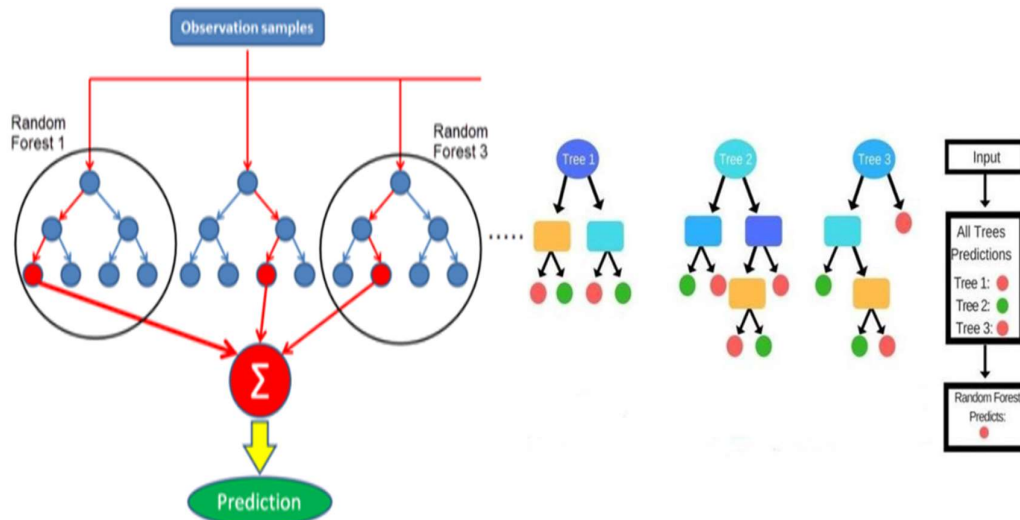
Trong đó  $I(p, i)$  gọi là chỉ số thu được;  $I$  kết quả chọn (*Information Gain*) giá trị  $p$  nhị phân (0,1) và  $n$  giá trị nhập thuộc bộ dữ liệu.

Chỉ số Entropy  $E$  được tính theo công thức:

$$\text{Entropy } E(A) = \sum_{i=1}^v \frac{p_i + n_i}{p+n} (I(p, n)) \quad (2)$$

Đây là chỉ số tính độ ổn định bộ số liệu. Tính chỉ số  $\text{Gain} = I(p, n) - E(A)$ .

Mô hình hóa qui trình xử lý thuật toán Random forest được thể hiện trên hình 3.



Hình 3. Các bước học (train) của thuật toán Random forest.

### 2.3. Các tham số vận dụng Class *sklearn.ensemble.RandomForestRegressor* trong Python

Cú pháp:

```
class sklearn.ensemble.RandomForestRegressor(n_estimators=100, *, criterion='square
d_error', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fracti
on_leaf=0.0, max_features='auto', max_leaf_nodes=None, min_impurity_decrease=0.0, bo
otstrap=True, oob_score=False, n_jobs=None, random_state=None, verbose=0, warm_start
=False, ccp_alpha=0.0, max_samples=None) [14–15].
```

Một số tham số chính:

- `n_estimators`: mặc định 100 : tổng số tham số (mỗi tham số tree) trong bộ dataset (forest).
- `criterion`: {"squared\_error", "mse", "absolute\_error", "poisson"}, default="squared\_error".

Chức năng để chọn một phương thức tính sai số dạng: Phương sai, sai số tuyệt đối, hay chỉ số poisson, mặc định chọn phương sai "squared\_error" [11–12].

- `max_depth`: chỉ số max số nhánh của bộ số liệu, đây là số nguyên, mặc định là None số nguyên int, default = None

- `min_samples_split`: Tham số min của chỉ số làm mẫu phân chia giữa số chọn và số không chọn, mặc định = 2

- `bootstrap`: kiểu bool, mặc định True, nếu chọn true: thì thuật toán lấy một số mẫu để xây dựng model, ngược lại false sẽ dùng toàn bộ dữ liệu để tạo model cho từng nhánh [8].

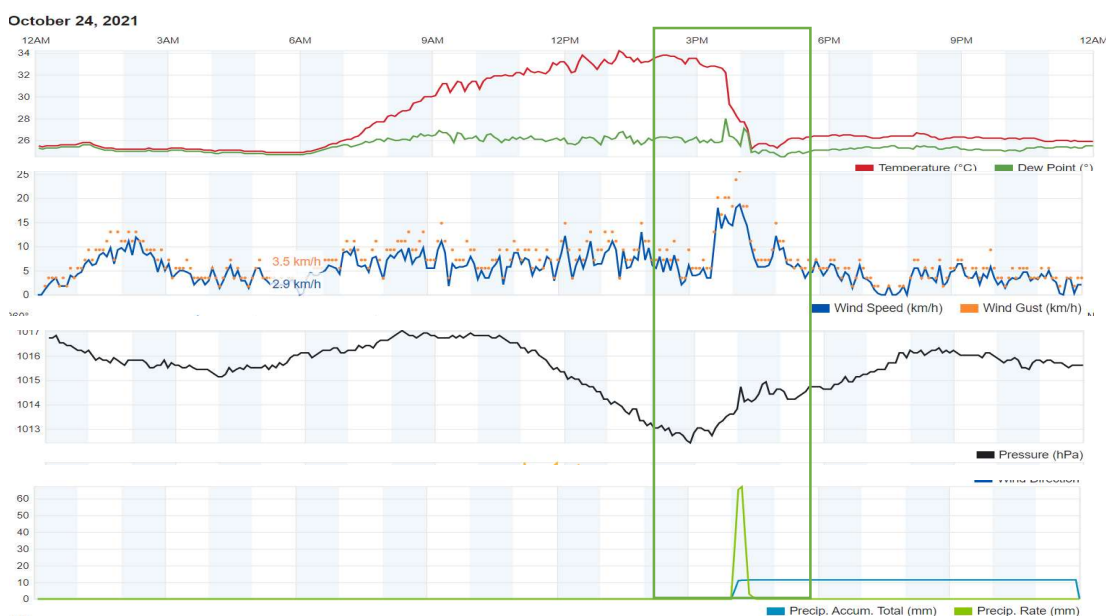
Bảng dataset mẫu được trình bày trong Bảng 2.

**Bảng 2.** Dữ liệu mẫu thu thập trạm Tăng Nhơn Phú A–TP. Thủ Đức.

| Thời gian        | Nhiệt độ (°C) | Độ ẩm | Tốc độ gió | Khí Áp | Mưa (t/f) |
|------------------|---------------|-------|------------|--------|-----------|
| 25/10/2020 15:35 | 31.7          | 70    | 8.5        | 29.879 | 0         |
| 25/10/2020 15:25 | 31.5          | 70    | 8.1        | 29.876 | 0         |
| 25/10/2020 15:15 | 31.7          | 69    | 3.6        | 29.876 | 0         |
| 25/10/2020 15:05 | 31.3          | 70    | 5.8        | 29.879 | 0         |
| 25/10/2020 14:55 | 31            | 71    | 2          | 29.882 | 0         |
| 25/10/2020 14:45 | 31            | 72    | 6.3        | 29.885 | 0         |
| 25/10/2020 14:35 | 31            | 74    | 2.9        | 29.885 | 0         |
| 09/10/2021 14:47 | 25.5          | 96    | 3.58       | 29.675 | 1         |
| 09/10/2021 14:37 | 26.5          | 94    | 7.61       | 29.686 | 1         |
| 01/10/2021 11:48 | 27.8          | 89    | 2.24       | 29.819 | 1         |
| 09/10/2021 14:47 | 25.5          | 96    | 3.58       | 29.675 | 1         |
| 09/10/2021 14:37 | 26.5          | 94    | 7.61       | 29.686 | 1         |

#### 2.4. Dữ liệu tính toán

Qua thực nghiệm nghiên cứu và phỏng vấn các chuyên gia khí tượng có kinh nghiệm nghiên cứu tiến hành chọn bộ số liệu thử nghiệm các yếu tố sau: Nhiệt độ không khí, độ ẩm không khí, tốc độ gió, khí áp và kết quả mưa (Hình 4).



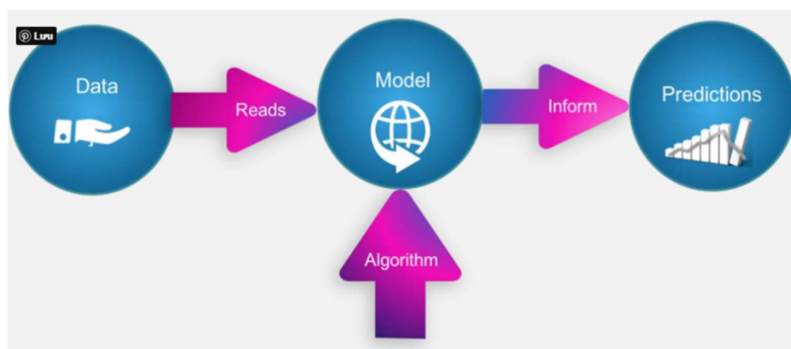
**Hình 4.** Biểu đồ diễn biến 4 yếu tố trước khi xảy ra sự kiện mưa.

Qua biểu đồ chúng ta so sánh sự biến thiên có xảy ra so với diễn biến bình thường không mưa trong bốn yếu tố được chọn để dự báo mưa.

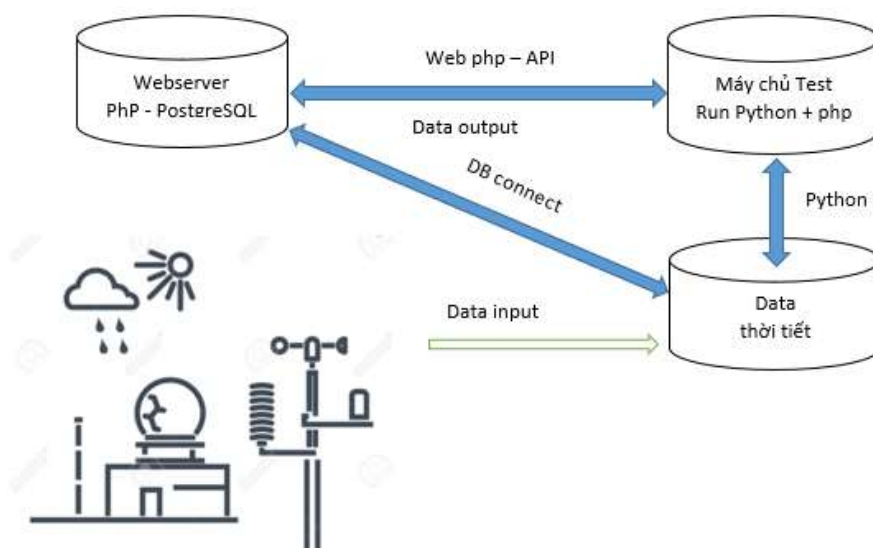
#### 2.5. Nguyên lý tính toán

Trong sơ đồ nguyên lý (Hình 5), nghiên cứu đã nhận diện các bước thực hiện, một về dữ liệu (data) phải có trước, lịch sử của dữ liệu càng nhiều càng tốt theo cấu trúc nhận dạng các yếu tố cần thiết, hai xây dựng các phương thức biên dịch thành các bảng mẫu so sánh (model) bảng mẫu có cơ chế tìm kiếm so sánh đặc biệt được mã hóa so sánh rất nhanh với dữ liệu đầu vào để trả về kết quả dự báo.



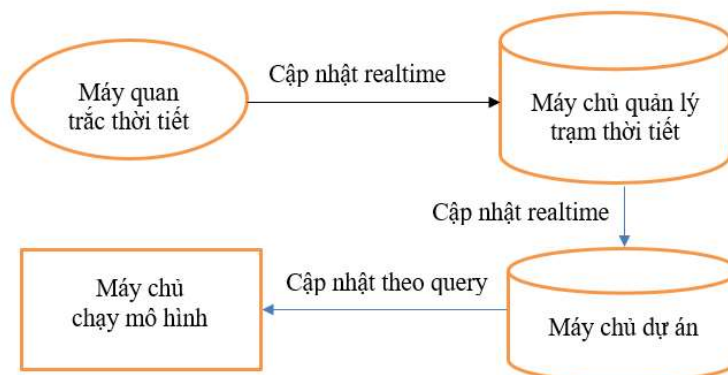


**Hình 5.** Nguyên lý hoạt động của Machine Learning trong việc dự báo mưa.



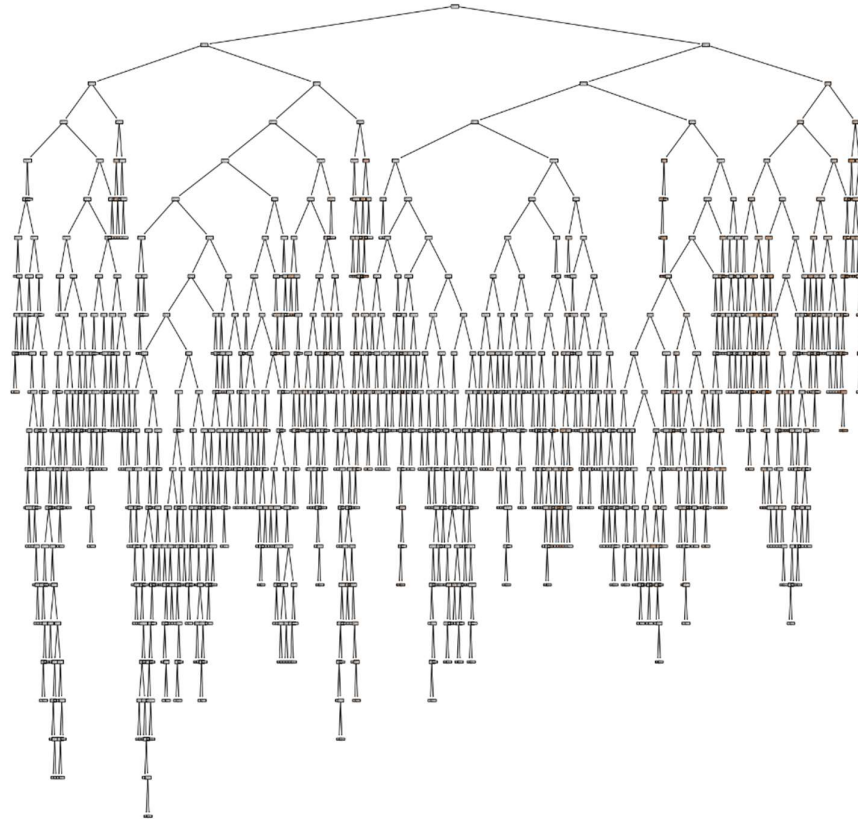
**Hình 6.** Sơ đồ khởi hoạt động của mô hình Machine Learning.

Quy trình vận hành hệ thống dự báo AI: Nghiên cứu tiến hành xây dựng hệ thống kết nối dữ liệu giữa các trạm thời tiết về trung tâm dữ liệu, từ trung tâm dữ liệu kết xuất ra các chuẩn định dạng dùng cho phần mềm học để tạo mô hình và truyền về hệ thống chạy dự báo theo sơ đồ hình 6, mỗi phần mềm có các chuẩn kết nối và truyền dữ liệu khác nhau vì vậy các hệ thống giao tiếp cần đến lập trình để đưa về các định dạng cần thiết phục vụ công tác lưu trữ, kết xuất, dự báo, trả kết quả cho phần mềm kế tiếp hoặc quay lại lưu trữ vào cơ sở dữ liệu.



**Hình 7.** Quy trình vận hành dữ liệu của hệ thống dự báo AI.

Hệ thống dự báo AI cần phải học dữ liệu liên tục và dự báo liên tục khi có dữ liệu mới phát sinh, đồng thời hệ thống xử lý song song cho nhiều điểm cùng một lúc. Cứ mỗi dòng dữ liệu cập nhật theo đúng thời gian qui định hệ thống sẽ tự kích hoạt ứng dụng dự báo chạy và xuất kết quả ghi vào database của các máy chủ cần nhận dữ liệu dự báo (Hình 7). Tất cả các công đoạn đều theo thời gian thực để đồng bộ dữ liệu giữa các điểm [19–21]. Qui trình xử lý Random forest được mô hình hóa bằng dữ liệu thực (Hình 8).



**Hình 8.** Mô hình tree được xây dựng khi chạy phần mềm dự báo.

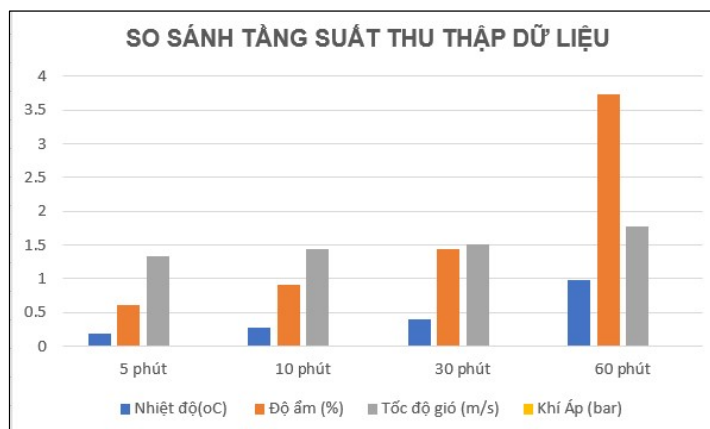
Đây là một cây diễn hình quá trình xử lý 1 node của hệ thống dự báo theo cách random forest, cứ mỗi nhiệt độ, độ ẩm, tốc độ gió, khí áp nhập vào hệ thống sẽ tra cứu theo mô hình đã biên dịch trước [17–18].

## 2.6. Tăng suất thu thập dữ liệu

Thống kê dữ liệu trong 2 ngày 23/10/2021 và 24/10/2021 những ngày có mưa tại trạm quan trắc Tăng Nhơn Phú A, tăng suất thấp nhất 5 phút /lần lấy mẫu. Nghiên cứu thống kê tần suất dữ liệu được thể hiện trong Bảng 3.

**Bảng 3.** Độ biến thiên trung bình giữa hai lần lấy mẫu các yếu tố.

| <b>Biến thiên</b> | <b>Nhiệt độ(°C)</b> | <b>Độ ẩm (%)</b> | <b>Tốc độ gió (m/s)</b> | <b>Khí Áp (bar)</b> |
|-------------------|---------------------|------------------|-------------------------|---------------------|
| 5 phút            | 0.189908            | 0.60551          | 1.3321                  | 0.00422             |
| 10 phút           | 0.281985            | 0.90625          | 1.4407                  | 0.00502             |
| 30 phút           | 0.409074            | 1.44444          | 1.5074                  | 0.00742             |
| 60 phút           | 0.982364            | 3.73921          | 1.7844                  | 0.01848             |



Hình 9. So sánh diễn biến giữa hai lần lấy mẫu các yếu tố.

Qua đồ thị chúng ta nhận thấy tăng suất thu thập dữ liệu 5 phút không khác biệt nhiều so với tăng suất 10 phút, nhưng có khác biệt lớn giữa 30 phút và 60 phút. Theo đồ thị 1 các diễn biến sắp có mưa xảy ra thì các yếu tố có biến thiên khác thường, khác với trung bình thì hệ thống học machine learning mới có thể nhận biết được điểm xuất hiện mưa hoặc không mưa. Tập dữ liệu huấn luyện phải đủ đảm bảo độ nhạy để máy nhận biết có biến thiên, nhưng vừa phải đảm bảo không trùng lặp (các dòng dữ liệu giống nhau) nếu độ khác biệt quá nhỏ xem như trùng lặp hệ thống sẽ xử lý rất chậm, tốn phí tài nguyên và thời gian khi chạy mô hình dự báo. Trong thí nghiệm chúng tôi chọn dữ liệu tăng suất 10 phút.

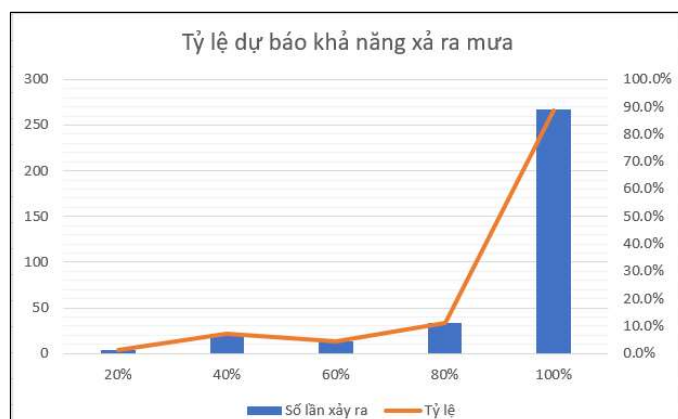
### 3. Kết quả và thảo luận

#### 3.1. Kết quả kiểm định mô hình dự báo mưa

Qua bảng thống kê kết quả trên chúng tôi nhận thấy: số lần dự báo khả năng mưa dưới 20% thực tế có xảy ra mưa chiếm 1.3%, số lần dự báo trên 60% khả năng mưa chiếm đến hơn 91% số lần xảy ra mưa thực tế, đặc biệt số lần dự báo 100% xảy ra mưa có đến 88.7% tại trạm Quan9 thời gian thực nghiệm từ tháng 10/2020 đến tháng 09/2021.

Bảng 4. Thống kê kết quả dự báo thực tế.

| Dự báo khả năng mưa | 20%  | 40%  | 60%  | 80%   | 100%  |      |
|---------------------|------|------|------|-------|-------|------|
| Số lần xảy ra       | 4    | 22   | 13   | 34    | 267   | 340  |
| Tỷ lệ               | 1,3% | 7,3% | 4,3% | 11,3% | 88,7% | 100% |



Hình 10. Kết quả dự báo mưa dùng thuật toán random forest trong thực tế trạm Quan9.



Hàm để xem accuary (độ chính xác) của model ứng với số k = 5.

In [14]: accuracy = rf\_model.score(X\_valid,Y\_valid); Out [15]: 0.9988913525498891

**Bảng 5.** Kết quả dự báo thời điểm trước khi có mưa khu vực TP. Thủ Đức.

| Trạm   | Số lần dự báo | Dự báo đúng | R <sup>2</sup> |
|--------|---------------|-------------|----------------|
| Thuduc | 2509          | 2238        | 0,89           |
| Quan2  | 2509          | 2388        | 0,95           |
| Quan9  | 2539          | 2501        | 0,95           |

**Bảng 6.** Kết quả dự báo thời điểm mưa khu vực TP. Thủ Đức.

| Trạm   | Số lần dự báo mưa | Số lần dự báo mưa đúng | R <sup>2</sup> |
|--------|-------------------|------------------------|----------------|
| Thuduc | 264               | 247                    | 0,93           |
| Quan2  | 118               | 92                     | 0,78           |
| Quan9  | 27                | 19                     | 0,7            |

Kết quả thể hiện số lần dự báo thời điểm trước khi có mưa và kết quả dự báo thời điểm có mưa cho thấy trạm thuduc, quan2, quan9 có kết quả dự báo chính xác cao nhất với hệ số tương quan lần lượt là 0,93, 0,78 và 0,7.

### 3.2. Kết quả dự báo mưa

Kết quả dự báo mưa tự động được lưu trữ trong máy chủ sau khi Run Python và php được hiển thị dưới dạng dữ liệu số và trên Web server (Hình 11–13).

Weather data log project

Chọn trạm:

| ID(ST)           | Thời gian           | Dữ liệu               | Dự báo | Thực tế |
|------------------|---------------------|-----------------------|--------|---------|
| 36948(binhchanh) | 2021-02-18 10:43:43 | 26.7-61.0-3.36-30.002 | 0      | 0       |
| 36943(binhchanh) | 2021-02-18 10:33:40 | 26.4-63.0-2.68-30.002 | 0      | 0       |
| 36938(binhchanh) | 2021-02-18 10:23:45 | 26.4-63.0-6.49-30.002 | 0      | 0       |
| 36933(binhchanh) | 2021-02-18 10:13:43 | 26.8-60.0-4.47-30.011 | 0      | 0       |
| 36928(binhchanh) | 2021-02-18 10:03:41 | 25.9-63.0-4.47-30.014 | 0      | 0       |
| 36923(binhchanh) | 2021-02-18 09:53:40 | 25.5-63.0-1.12-30.011 | 0      | 0       |
| 36918(binhchanh) | 2021-02-18 09:43:40 | 26.1-62.0-1.79-30.014 | 0      | 0       |
| 36913(binhchanh) | 2021-02-18 09:33:42 | 25.6-60.0-0.22-30.002 | 0      | 0       |
| 36908(binhchanh) | 2021-02-18 09:23:43 | 25.5-62.0-2.46-30.014 | 0      | 0       |
| 36903(binhchanh) | 2021-02-18 09:13:41 | 24.8-63.0-1.34-30.023 | 0      | 0       |
| 36898(binhchanh) | 2021-02-18 09:03:40 | 24.1-65.0-1.79-30.008 | 0      | 0       |
| 36893(binhchanh) | 2021-02-18 08:53:42 | 25.1-61.0-1.57-30.005 | 0      | 0       |
| 36888(binhchanh) | 2021-02-18 08:43:41 | 23.9-65.0-0.00-30.002 | 0      | 0       |
| 36883(binhchanh) | 2021-02-18 08:33:42 | 23.8-67.0-1.34-30.005 | 0      | 0       |
| 36878(binhchanh) | 2021-02-18 08:23:42 | 23.1-68.0-1.34-29.991 | 0      | 0       |
| 36873(binhchanh) | 2021-02-18 08:13:42 | 22.8-70.0-2.46-29.994 | 0      | 0       |
| 36868(binhchanh) | 2021-02-18 08:03:41 | 22.0-71.0-0.45-29.997 | 0      | 0       |
| 36863(binhchanh) | 2021-02-18 07:53:43 | 21.5-69.0-3.13-29.985 | 0      | 0       |
| 36858(binhchanh) | 2021-02-18 07:43:42 | 21.4-74.0-2.24-29.985 | 0      | 0       |
| 36853(binhchanh) | 2021-02-18 07:33:43 | 21.0-74.0-3.36-29.970 | 0      | 0       |
| 36848(binhchanh) | 2021-02-18 07:23:44 | 21.1-76.0-1.34-29.973 | 0      | 0       |
| 36843(binhchanh) | 2021-02-18 07:13:42 | 20.9-79.0-0.22-29.964 | 0      | 0       |
| 36838(binhchanh) | 2021-02-18 07:03:45 | 20.4-82.0-0.00-29.961 | 0      | 0       |
| 36833(binhchanh) | 2021-02-18 06:53:43 | 20.2-83.0-1.12-29.958 | 0      | 0       |

Weather data log project

Chọn trạm:

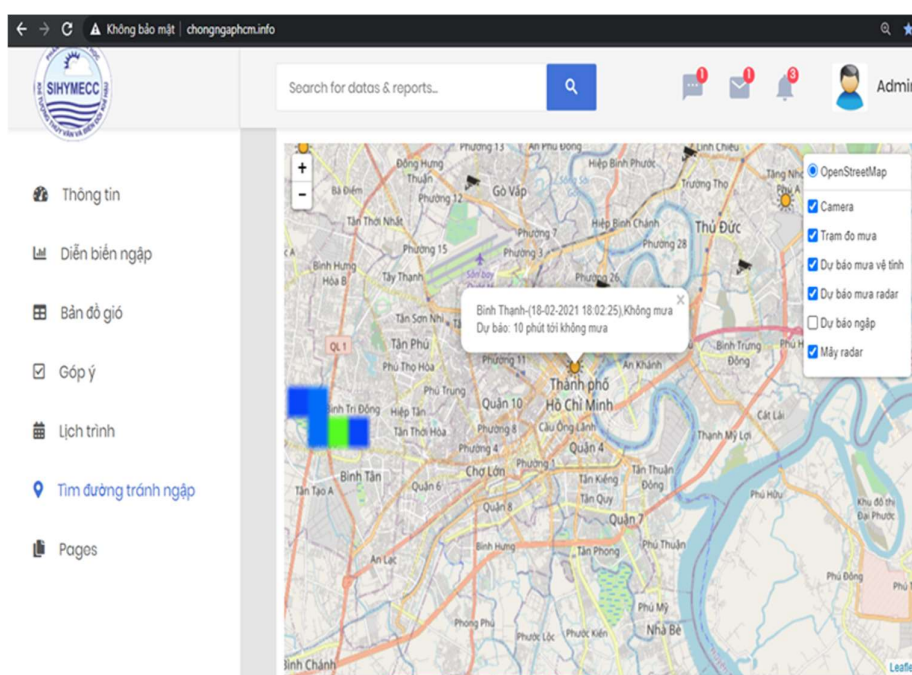
| ID(ST)       | Thời gian           | Dữ liệu               | Dự báo | Thực tế |
|--------------|---------------------|-----------------------|--------|---------|
| 36946(quan2) | 2021-02-18 10:43:40 | 28.3-56.0-2.91-30.008 | 0      | 0       |
| 36941(quan2) | 2021-02-18 10:33:37 | 27.9-58.0-1.12-30.011 | 0      | 0       |
| 36936(quan2) | 2021-02-18 10:23:39 | 27.3-61.0-0.67-30.017 | 0      | 0       |
| 36931(quan2) | 2021-02-18 10:13:40 | 27.1-62.0-1.57-30.020 | 0      | 0       |
| 36926(quan2) | 2021-02-18 10:03:39 | 27.7-60.0-0.22-30.026 | 0      | 0       |
| 36921(quan2) | 2021-02-18 09:53:38 | 26.8-64.0-0.89-30.020 | 0      | 0       |
| 36916(quan2) | 2021-02-18 09:43:38 | 26.4-65.0-0.22-30.017 | 0      | 0       |
| 36911(quan2) | 2021-02-18 09:33:39 | 26.5-63.0-2.91-30.020 | 0      | 0       |
| 36906(quan2) | 2021-02-18 09:23:38 | 26.5-64.0-1.12-30.017 | 0      | 0       |
| 36901(quan2) | 2021-02-18 09:13:38 | 26.4-64.0-0.22-30.017 | 0      | 0       |
| 36896(quan2) | 2021-02-18 09:03:38 | 25.6-67.0-1.79-30.026 | 0      | 0       |
| 36891(quan2) | 2021-02-18 08:53:39 | 25.5-70.0-0.22-30.014 | 0      | 0       |
| 36886(quan2) | 2021-02-18 08:43:38 | 25.3-70.0-0.89-30.017 | 0      | 0       |
| 36881(quan2) | 2021-02-18 08:33:38 | 24.2-67.0-0.22-30.011 | 0      | 0       |
| 36876(quan2) | 2021-02-18 08:23:39 | 23.4-69.0-0.00-30.008 | 0      | 0       |
| 36871(quan2) | 2021-02-18 08:13:38 | 23.1-69.0-0.00-30.000 | 0      | 0       |
| 36866(quan2) | 2021-02-18 08:03:38 | 22.7-70.0-0.00-30.005 | 0      | 0       |
| 36861(quan2) | 2021-02-18 07:53:40 | 22.3-72.0-0.00-30.000 | 0      | 0       |
| 36856(quan2) | 2021-02-18 07:43:39 | 22.2-73.0-0.00-30.002 | 0      | 0       |
| 36851(quan2) | 2021-02-18 07:33:40 | 21.8-73.0-0.00-29.988 | 0      | 0       |
| 36846(quan2) | 2021-02-18 07:23:40 | 21.8-74.0-0.00-29.991 | 0      | 0       |
| 36841(quan2) | 2021-02-18 07:13:40 | 21.7-75.0-0.45-29.976 | 0      | 0       |
| 36836(quan2) | 2021-02-18 07:03:40 | 21.8-75.0-0.00-29.979 | 0      | 0       |
| 36831(quan2) | 2021-02-18 06:53:40 | 21.9-72.0-0.67-29.973 | 0      | 0       |

**Hình 11.** Kết quả dự báo mưa tự động được lưu trữ trên máy chủ sever, với kết quả dự báo và thực tế.

Ghi chú: 0 là không có khả năng mưa, 1 là có khả năng gây mưa

|  |
|--|
| 218:P. LINH ĐÔNG;01/10/2020 18:00:00;0 |
| 218:P. LINH ĐÔNG;01/10/2020 17:00:00;0 |
| 218:P. LINH ĐÔNG;01/10/2020 16:00:00;0 |
| 218:P. LINH ĐÔNG;01/10/2020 15:00:00;0 |
| 218:P. LINH ĐÔNG;01/10/2020 14:00:00;0 |
| 218:P. LINH ĐÔNG;01/10/2020 13:00:00;0 |
| 218:P. LINH ĐÔNG;01/10/2020 12:00:00;0 |
| 218:P. LINH ĐÔNG;01/10/2020 11:00:00;0 |
| 218:P. LINH ĐÔNG;01/10/2020 10:00:00;0 |
| 218:P. LINH ĐÔNG;01/10/2020 09:00:00;0 |
| 218:P. LINH ĐÔNG;01/10/2020 08:00:00;0 |
| 218:P. LINH ĐÔNG;01/10/2020 07:00:00;0 |
| 218:P. LINH ĐÔNG;01/10/2020 06:00:00;0 |
| 218:P. LINH ĐÔNG;01/10/2020 05:00:00;0 |
| 12:P. AN PHU;24/12/2020 18:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 20:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 21:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 22:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 17:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 23:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 13:00:00;0.30  |
| 12:P. AN PHU;24/12/2020 08:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 14:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 09:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 15:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 10:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 16:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 11:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 12:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 07:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 02:00:00;0.10  |
| 12:P. AN PHU;24/12/2020 03:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 04:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 05:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 06:00:00;0.00  |
| 12:P. AN PHU;24/12/2020 01:00:00;0.00  |

Hình 12. Kết quả dự báo mưa từ các trạm đo mưa.



Hình 13. Kết quả dự báo mưa tự động được hiển thị trên WebGIS.

#### 4. Kết luận

Về việc xây dựng mô hình dự báo mưa đã hoàn thiện phần xây dựng được mô hình dự báo mưa với các bước đầu là thu thập các nguồn số liệu từ các trạm đo mưa tự động, số hóa và xử lý thành công các dữ liệu thời tiết và hiện thị trên công cụ WebGIS của nghiên cứu, tổ chức các buổi hội thảo tham khảo các ý kiến từ các chuyên gia về dự báo mưa, khí tượng, để có thể kết hợp phương pháp dự báo mưa truyền thống với công nghệ trí tuệ nhân tạo AI. Việc dự báo mưa đã cho được kết quả dự báo khá tốt về có mưa hay không mưa thông qua kết quả dự báo mưa tự động được hiện thị trên nền tảng WebGIS, cùng với đó là kết hợp với các trang dự báo uy tín, để tiến hành hiệu chỉnh và kiểm định kết quả trước khi làm cơ sở đầu vào cho mô hình dự báo ngập. Tuy nhiên, mô hình dự báo mưa tiếp tục được hoàn thiện về dự báo ra giá trị định lượng mưa và không gian mưa ở đâu nhằm đưa ra được đầu vào chính xác cho mô hình dự báo ngập.

**Đóng góp của tác giả:** Xây dựng ý tưởng nghiên cứu: P.T.L., L.V.P., T.T.H., L.H.D., N.P.Đ.; Lựa chọn phương pháp nghiên cứu: L.V.P., P.T.L., T.T.H.; Xử lý số liệu: N.P.Đ., L.V.P.; Viết bản thảo bài báo: P.T.L., N.P.Đ.; Chỉnh sửa bài báo: P.T.L., L.V.P., T.T.H.

**Lời cảm ơn:** Nghiên cứu này được xây dựng dựa trên kết quả nghiên cứu của nhiệm vụ: “Xây dựng mô hình dự báo, cảnh báo và quản lý ngập cho đô thị thông minh tại Thành phố Hồ Chí Minh”.

**Lời cam đoan:** Tập thể tác giả cam đoan bài báo này là công trình nghiên cứu của tập thể tác giả; không có sự tranh chấp lợi ích trong nhóm tác giả.

#### Tài liệu tham khảo

1. <https://techcrunch.com/2020/09/01/googles-ai-powered-flood-alerts-now-cover-all-of-india-and-parts-of-bangladesh/> [Accessed: 10 Oct 2021].
2. Loi, N. K., H.T.; Duong, T.Q.; Nguyen, L.D.; Vo, T.Q.N.; Tran, N.T.; Dang, P.D.N.; Nguyen, L.D.; Dang, C.K.; Nguyen, L.K. Development of a Spatial Decision Support System for Real-Time Flood Early Warning in the Vu Gia-Thu Bon River Basin, Quang Nam Province, Vietnam. *Sensors* **2020**, *20*(6), 1667. <https://doi.org/10.3390/s20061667>.
3. Keoduangsine, S.; Robert, R.; Gardner-Stephen, P. A Review of Flood Warning Systems in Developed and Developing Countries. Commission International Journal of Future Computer and Communication, 2014.
4. Flood early warning system. Design, implementation and computational modules - International Conference on Computational Science, ICCS, 2011.
5. Hashi, A.O.; Abdirahman, A.A.; Elmi, M.A.; Hashi, S.Z.M.; Rodriguez, O.E.R. A Real-Time Flood Detection System Based on Machine Learning Algorithms with Emphasis on Deep Learning. *Int. J. Eng. Trends Technol.* **2021**, *69*(5), 249–256.
6. Phân Viện Khí tượng Thủy văn và Biến đổi khí hậu. Nghiên cứu xây dựng cơ sở dữ liệu các đặc trưng khí tượng thủy văn phục vụ phòng chống ngập úng khu vực TP. HCM, 2008.
7. Phân viện Khí tượng Thủy văn và Biến đổi khí hậu. Xây dựng, cập nhật kế hoạch hành động ứng phó với BĐKH giai đoạn 2021-2030, tầm nhìn đến 2050 trên địa bàn TPHCM. Dự án Sở TNMT TP.HCM, 2020.
8. <https://www.rxdatscience.com/blog/machine-learning-for-pharma-using-random-forest>.
9. Sejnowski, T.J.; Tài, H.H. Deep Learning - Cuộc Cách Mạng Học Sâu. NXB Công Thương, 2018.
10. Chollet, F. Deep Learning with Python, Public: Manning Publications, 2017.
11. <https://medium.com/machine-learning-researcher/random-forest-algorithm-in-machine-learning-b15906b33157>.
12. Thúc, N.Đ. Trí Tuệ Nhân Tạo Mạng Noron - Phương Pháp Và Ứng Dụng. NXB Giáo Dục, 2002.
13. Keoduangsine, S.; Robert Goodwin, R. An Appropriate Flood Warning System in the Context of Developing Countries. *Int. J. Innovation Manage. Technol.* **2012**, *3*(3), 213–2016.
14. Flood Early Warning System Using Cognitive Artificial Intelligence. The Design of AWLR Sensor, 2018 International Conference on Information Technology Systems and Innovation (ICITSI), Bandung, Indonesia, 2019.
15. Mohd Sabre, M.S.; Abdullah, S.S.; Amrul Faruq, A. Flood Warning and Monitoring System Utilizing Internet of Things Technology. Kinetik Game Technology Information System Computer Network Computing Electronics and Control, 2019.

16. Rodríguez, S.O.; Wang, L.P.; Thraves, L.; Johnston, A.; Onof, C. Surface water flood <https://www.nature.com/articles/s41598-019-56452-5>/Short-term rainfall forecast model based on the improved BP–NN algorithm.
17. Mather, B. Machine Learning in Python: Hands on Machine Learning with Python Tools, Concepts and Techniques. Russell Newton, 2020.
18. Building an Intelligent Hydroinformatics Integration Platform for Regional Flood Inundation Warning Systems. Online available: <https://www.mdpi.com/2073-4441/11/1/9/htm>.
19. warnings in England: overview, assessment and recommendations based on survey responses and workshops. *J. Flood Risk Manage.* **2015**, *11*, S211–S221. <https://doi.org/10.1111/jfr3.12195>.
20. al Makhfi Indar, J. An intelligent Moroccan flood prediction and early warning system. Software Project, 2019.
21. Gastn, C.H. Building RESTful Python Web Services. Packt Publishing Ltd., 2016.

## Research on pilot AI rain forecasting technology in Ho Chi Minh City

Pham Thanh Long<sup>1</sup>, Le Van Phan<sup>2</sup>, Nguyen Phuong Dong<sup>1</sup>, Le Hong Duong<sup>3</sup>, Tran Tuan Hoang<sup>1</sup>

<sup>1</sup> Sub-Institute of Hydrometeorology and Climate Change; longpham.sihymete@gmail.com; hoangkttv@gmail.com; nguyenphuongdongkttv@gmail.com

<sup>2</sup> Nong Lam University; lvphan@hcmuaf.edu.vn

<sup>3</sup> Department of Southern Environmental Protection; hongduong.tcmt@gmail.com

**Abstract:** AI technology has been applied in weather forecasting in the developed countries are very popular and more accurately. In recent years, the monitoring data system updated in real-time of HydroMet stations has developed very strong, the data sources are large and continuous that is great advantage in AI rain forecast, pilot in Ho Chi Minh City. The factors of air temperature, air humidity, wind speed, barometric pressure at the monitoring point are input data for the AI rain forecasting system. Based on real-time weather monitoring stations, computer-based AI software processing technology is very fast, the process of learning data and forecasting new results is very fast and convenient. We have conducted testing and forecasting with quite high accuracy. The results showing the number of times of forecasting the time before the rain and the forecasting results of the time of rain show that thuduc, quan2, quan9 stations have the highest accurate forecast results with correlation coefficients of 0.93, respectively 0.78 and 0.7.

**Keywords:** AI rain forecast; Artificial intelligence; Big data; Data science.