

HUST

ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.



ĐẠI HỌC
BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Recurrent Neural Networks

ONE LOVE. ONE FUTURE.

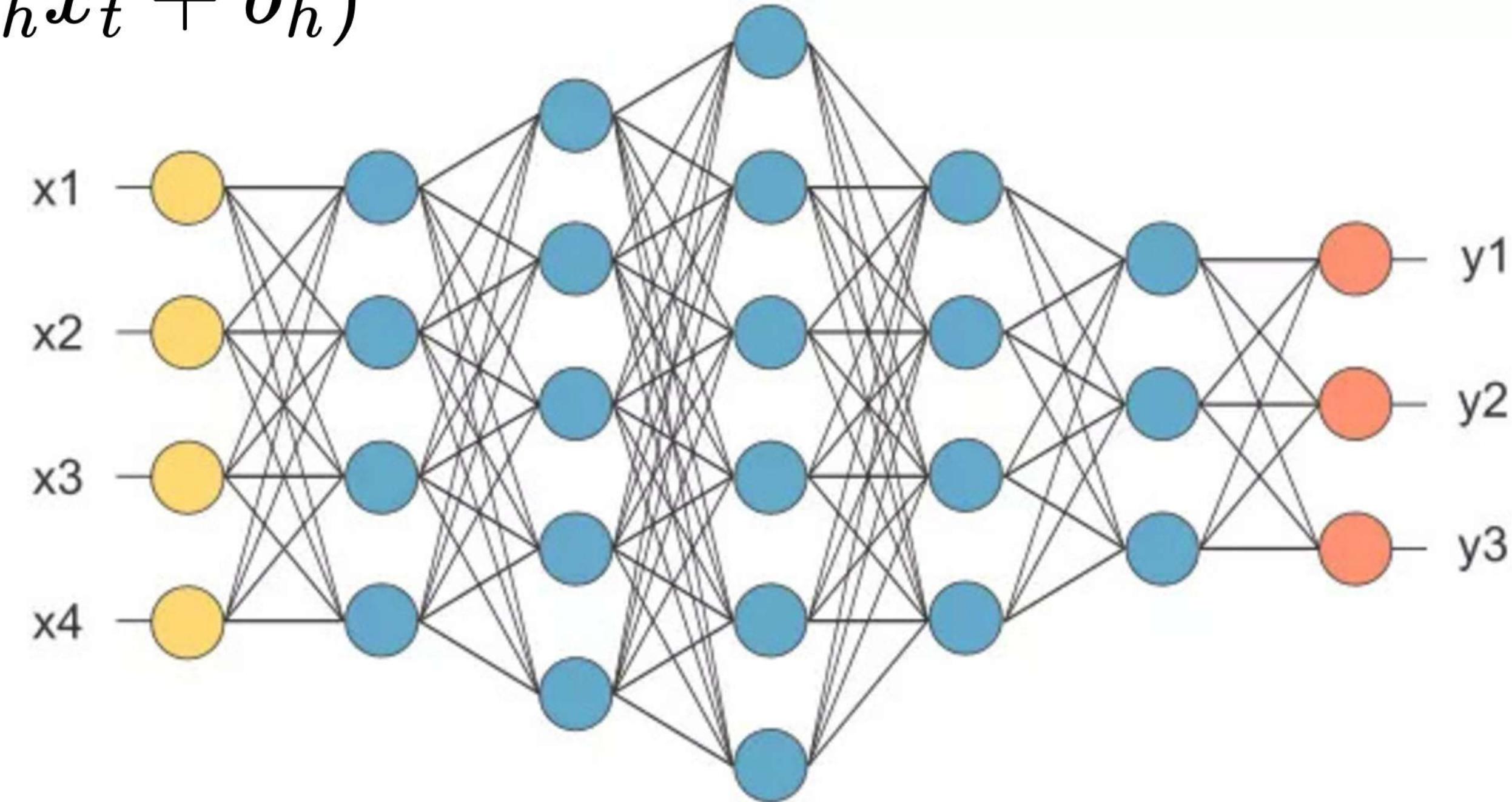
Nội dung chính

- Recurrent Neural Networks (RNN)
- Long Short-Term Memory (LSTM)
- Gated Recurrent Units (GRU)
- Ứng dụng bài toán thực tế



Multilayer Perceptron

$$h_t = f(W_h x_t + b_h)$$



Recurrent Neural Networks

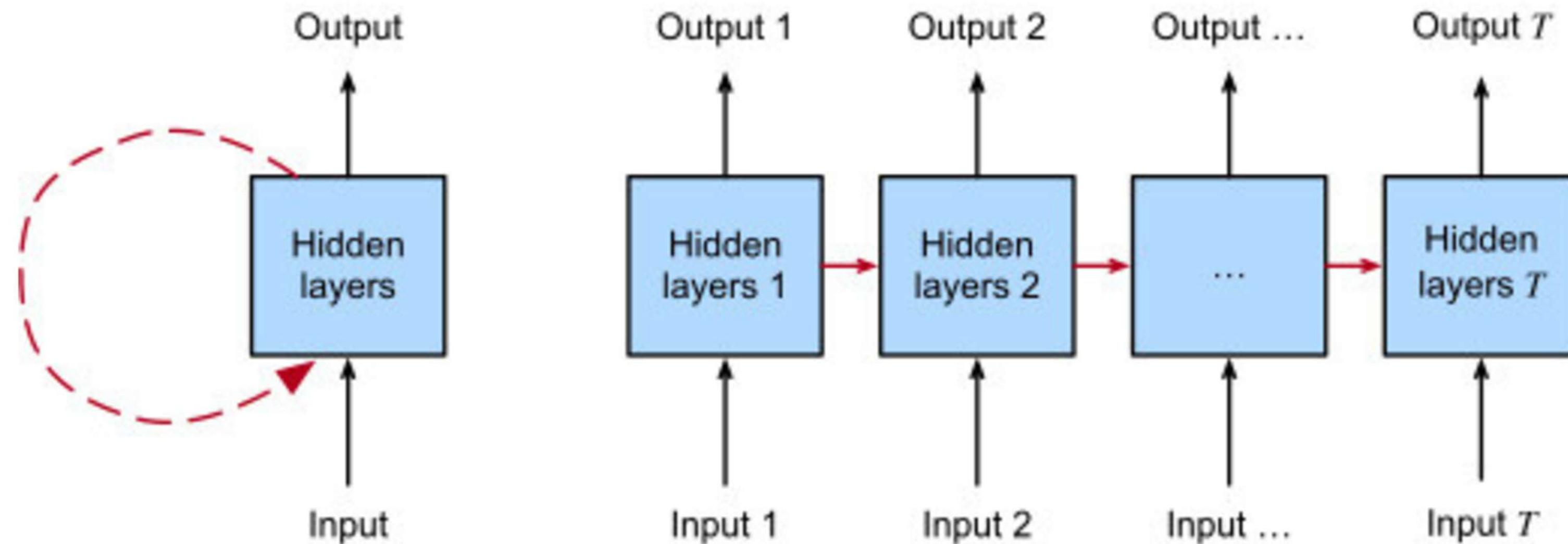
- Recurrent Neural Networks (RNN) cho phép mạng “ghi nhớ trạng thái trước đó” thông qua hidden state được truyền từ bước này sang bước tiếp theo. Mạng giờ đây có bộ nhớ tạm thời (short-term memory)

$$h_t = f(W_{ih}x_t + b_{ih} + W_{hh}h_{t-1} + b_{hh})$$

$$= f\left(\begin{bmatrix} W_{ih} \\ W_{hh} \end{bmatrix} [x_t \quad h_{t-1}] + (b_{ih} + b_{hh})\right)$$



Recurrent Neural Networks



Backpropagation through Time

- Forward propagation

$$h_t = f(x_t, h_{t-1}, w_h) \quad o_t = g(h_t, w_o)$$

$$L(x_1, \dots, x_T, y_1, \dots, y_T, w_h, w_o) = \frac{1}{T} \sum_{t=1}^T l(y_t, o_t)$$

- Backward propagation

$$\begin{aligned} \frac{dL}{dw_h} &= \frac{1}{T} \sum_{t=1}^T \frac{dl(y_t, o_t)}{dw_h} = \frac{1}{T} \sum_{t=1}^T \frac{dl(y_t, o_t)}{do_t} \frac{dg(h_t, w_o)}{dw_h} \\ &= \frac{1}{T} \sum_{t=1}^T \frac{dl(y_t, o_t)}{do_t} \frac{dg(h_t, w_o)}{dh_t} \frac{dh_t}{dw_h} \end{aligned}$$

Backpropagation through Time

- Forward propagation

$$h_t = f(x_t, h_{t-1}, w_h) \quad o_t = g(h_t, w_o)$$

$$L(x_1, \dots, x_T, y_1, \dots, y_T, w_h, w_o) = \frac{1}{T} \sum_{t=1}^T l(y_t, o_t)$$

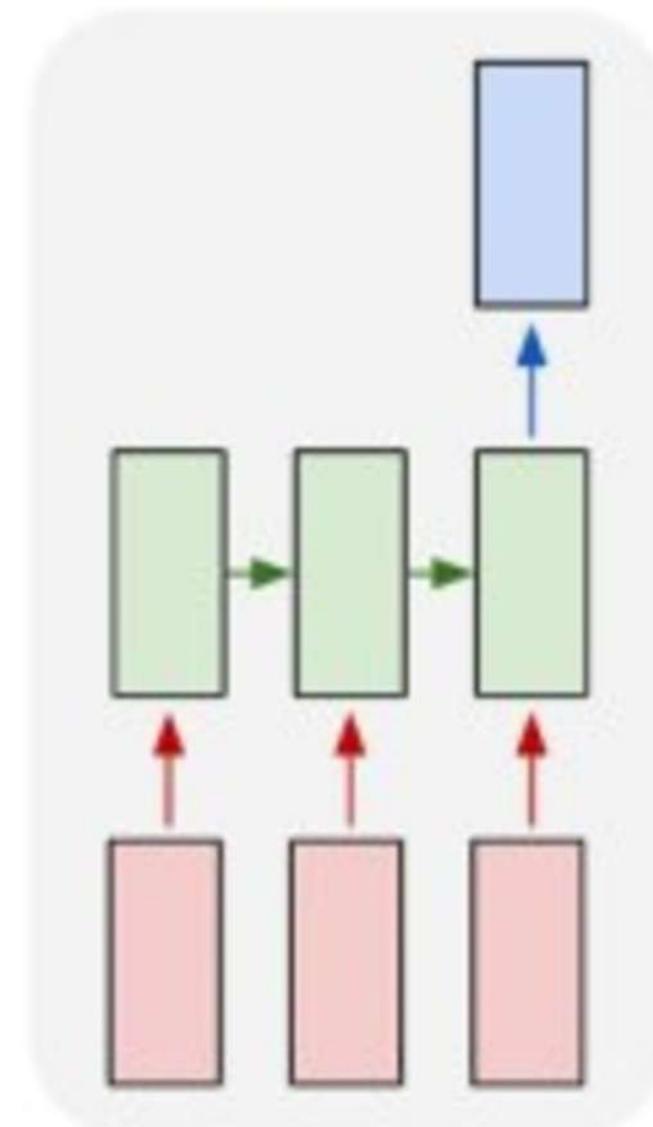
- Backward propagation

$$\frac{dh_t}{dw_h} = \frac{df(x_t, h_{t-1}, w_h)}{dw_h} + \frac{df(x_t, h_{t-1}, w_h)}{dh_{t-1}} \frac{dh_{t-1}}{dw_h}$$

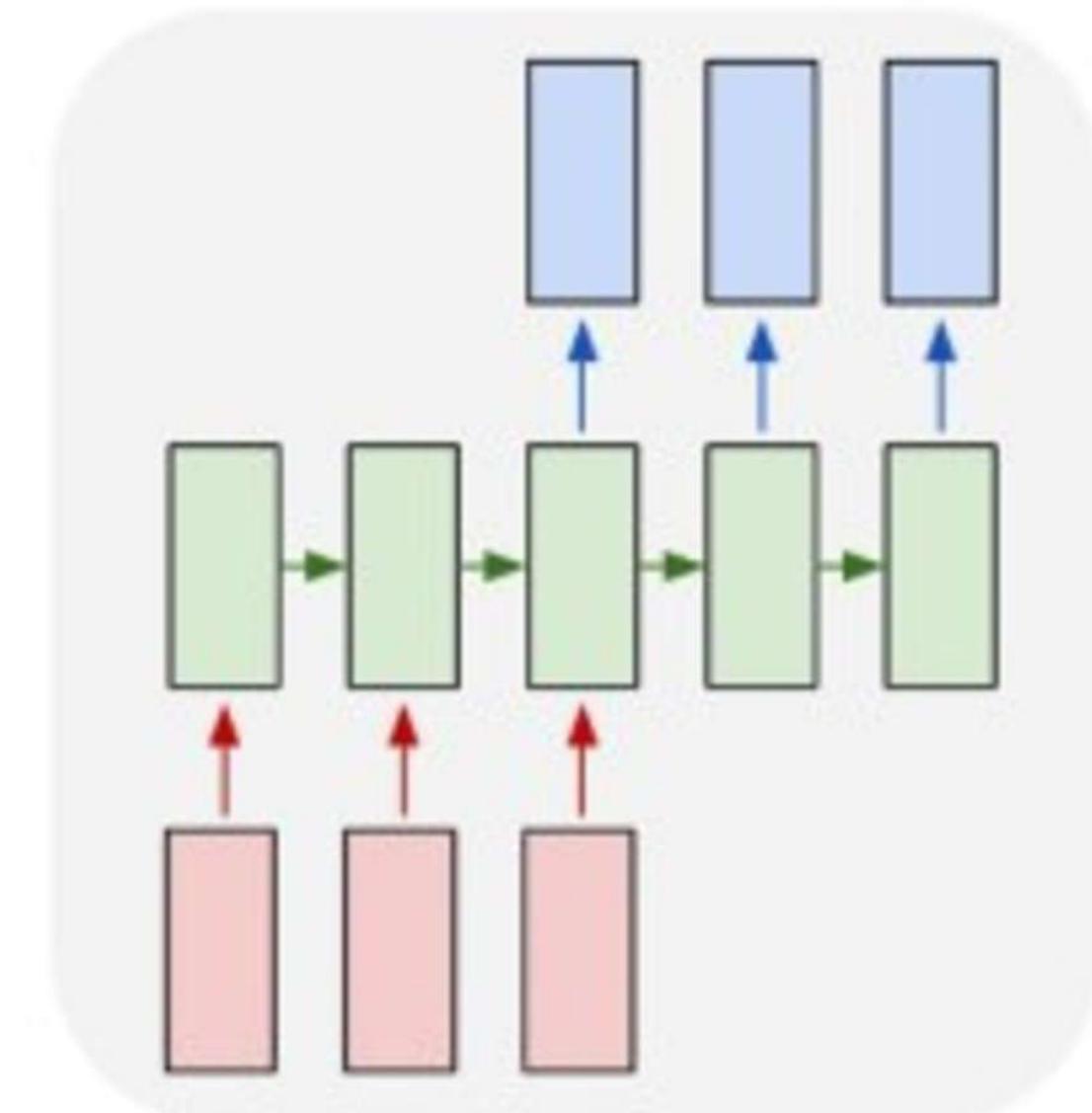
$$= \frac{df(x_t, h_{t-1}, w_h)}{dw_h} + \sum_{i=1}^{t-1} \left(\prod_{j=i+1}^t \frac{df(x_t, h_{j-1}, w_h)}{dh_{j-1}} \right) \frac{df(x_i, h_{i-1}, w_h)}{dw_h}$$

Recurrent Neural Networks

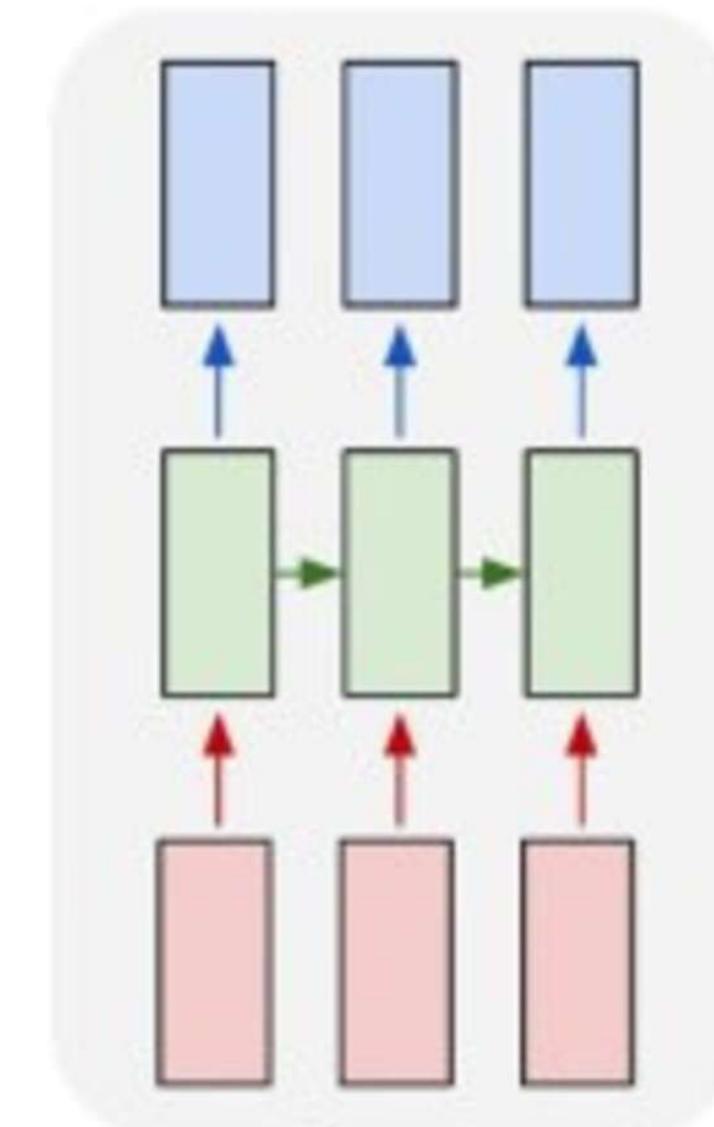
many to one



many to many



many to many



- Trong cấu trúc RNN, Sigmoid và Tanh được ưu tiên sử dụng hơn là ReLU:
 - ReLU giúp tránh hiện tượng vanishing gradient, nhưng dễ tạo ra hiện tượng exploding gradient khi số bước thời gian lớn
 - Sigmoid và Tanh giới hạn đạo hàm giúp không bị exploding gradient

Long Short-Term Memory

- Long Short-Term Memory (LSTM) được thiết kế để tránh hiện tượng vanishing gradient ở RNN bằng cách:
 - Tạo thêm một đường truyền mà gradient gần như không bị suy giảm đóng vai trò như một trí nhớ dài hạn (long-term memory)
 - Điều chỉnh lượng thông tin cần lưu trữ, lãng quên và thêm vào ở mỗi bước thời gian

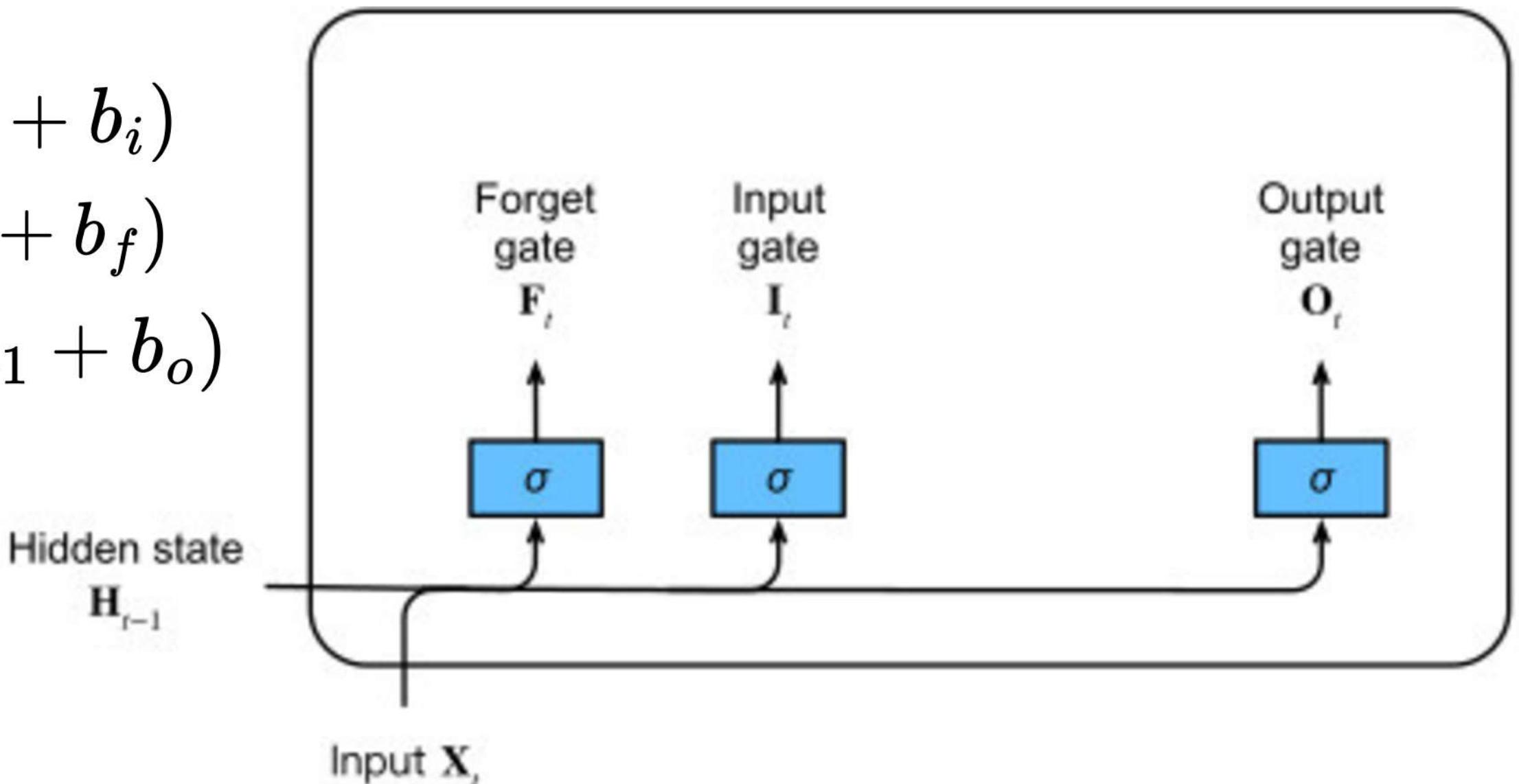


Long Short-Term Memory

$$I_t = \sigma(W_{xi}X_t + W_{hi}H_{t-1} + b_i)$$

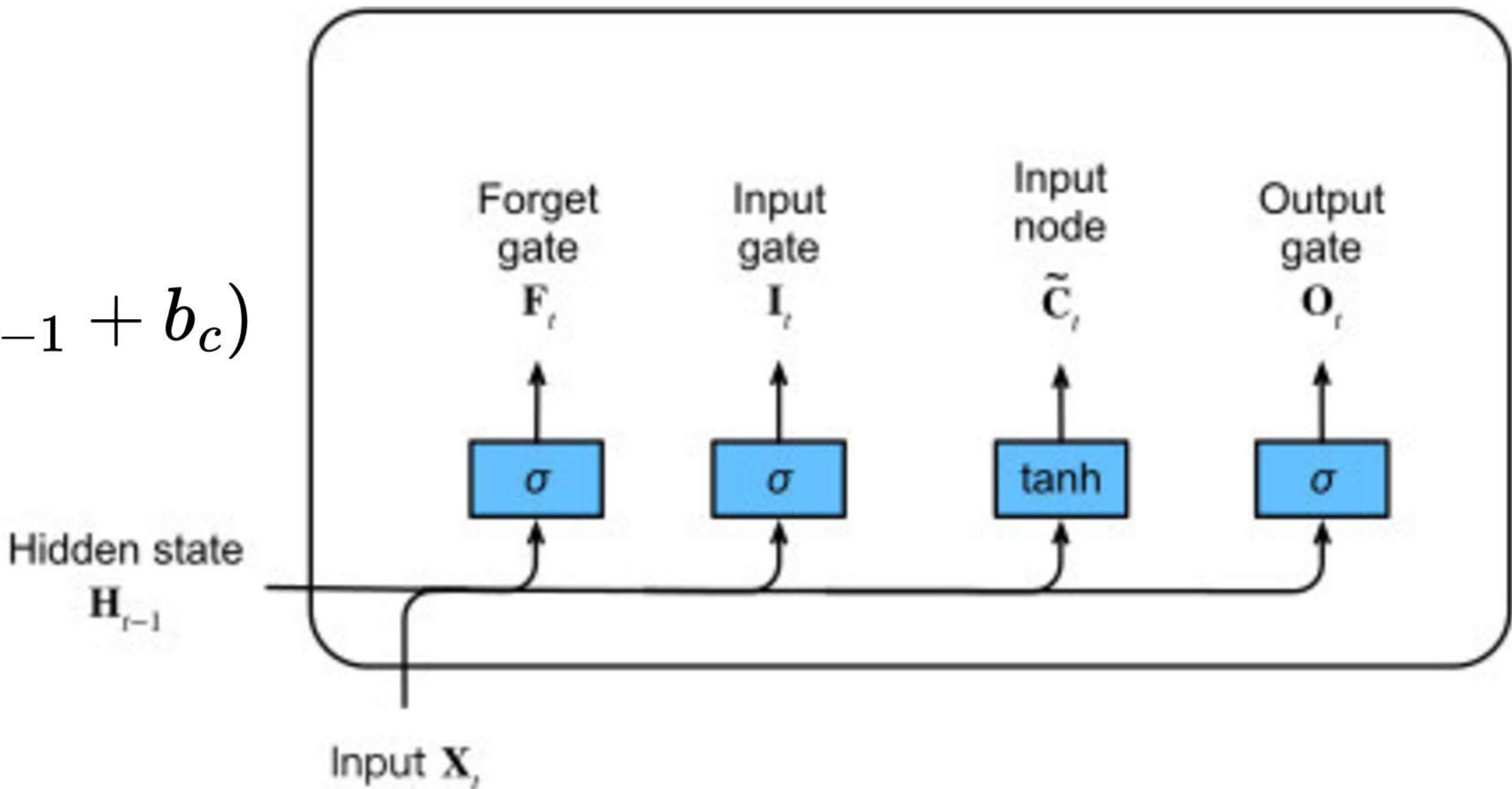
$$F_t = \sigma(W_{xf}X_t + W_{hf}H_{t-1} + b_f)$$

$$O_t = \sigma(W_{xo}X_t + W_{ho}H_{t-1} + b_o)$$



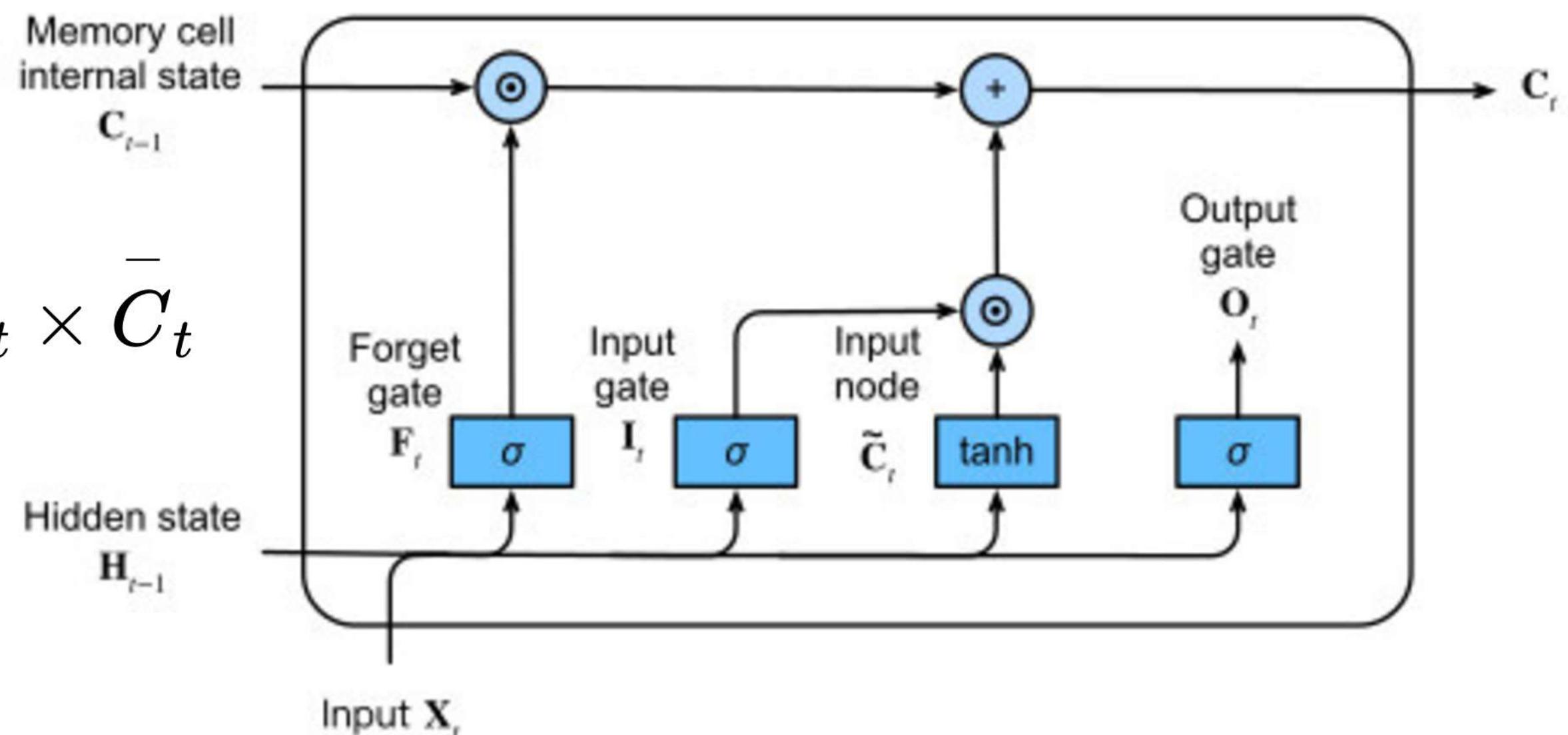
Long Short-Term Memory

$$\bar{C}_t = \tanh(W_{xc}X_t + W_{hc}H_{t-1} + b_c)$$



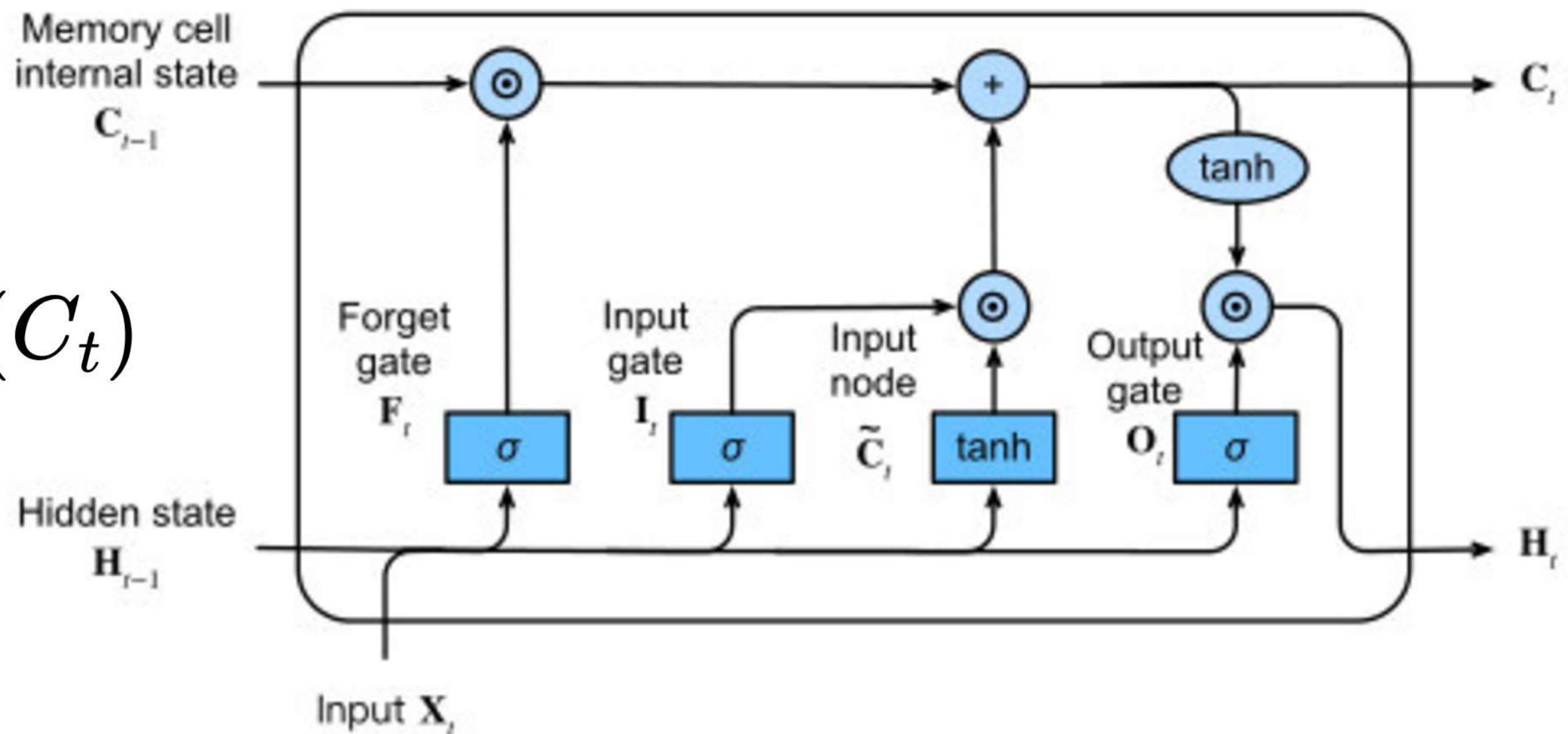
Long Short-Term Memory

$$C_t = F_t \times C_{t-1} + I_t \times \tilde{C}_t$$



Long Short-Term Memory

$$H_t = O_t \times \tanh(C_t)$$



Gated Recurrent Units

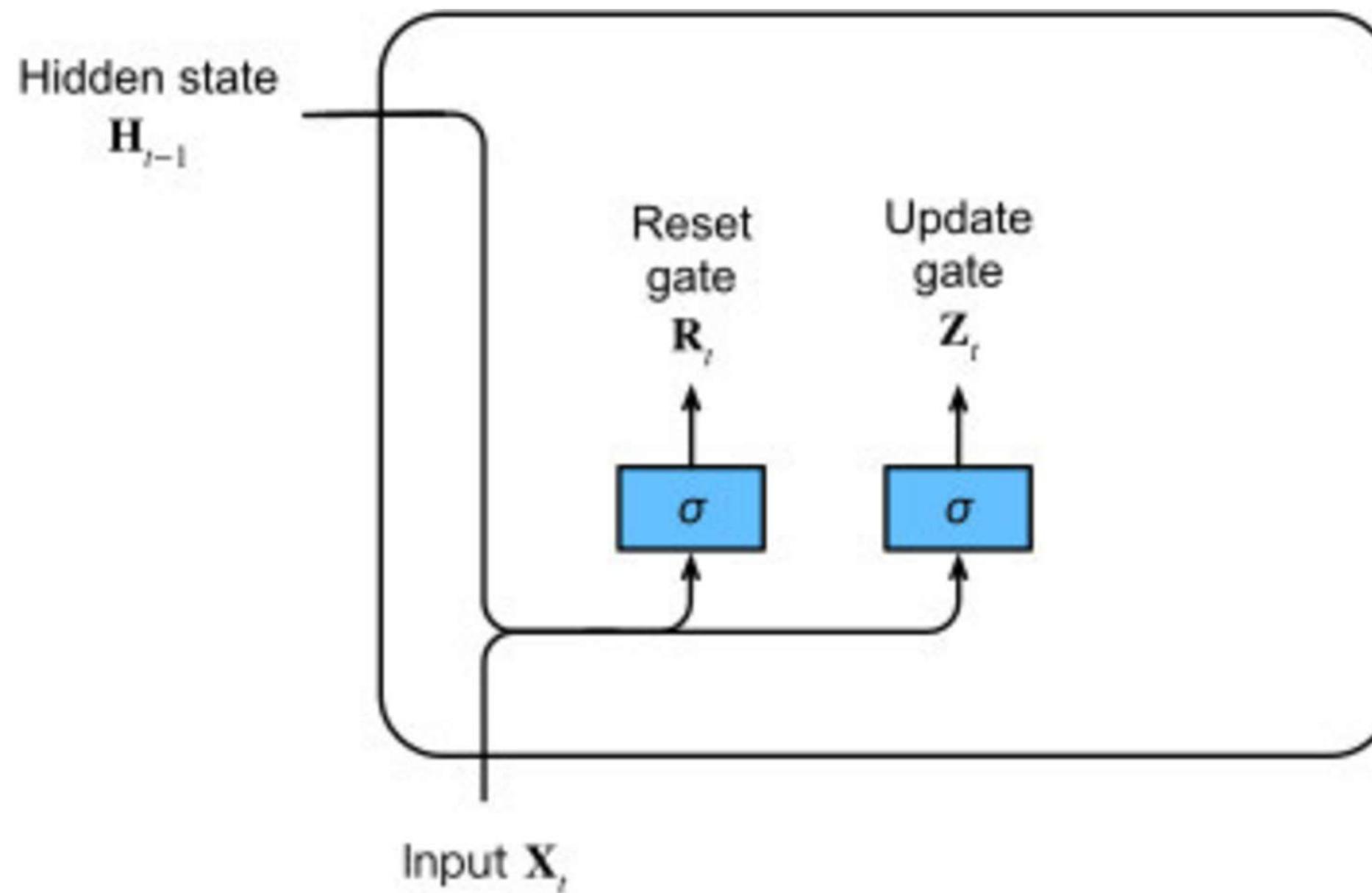
- Gated Recurrent Units (GRU) là phiên bản tinh gọn hơn của LSTM khi gộp hidden state và cell state vào làm một
- GRU đơn giản hơn, huấn luyện nhanh hơn, nhưng vẫn đảm bảo hiệu quả không kém LSTM



Gated Recurrent Units

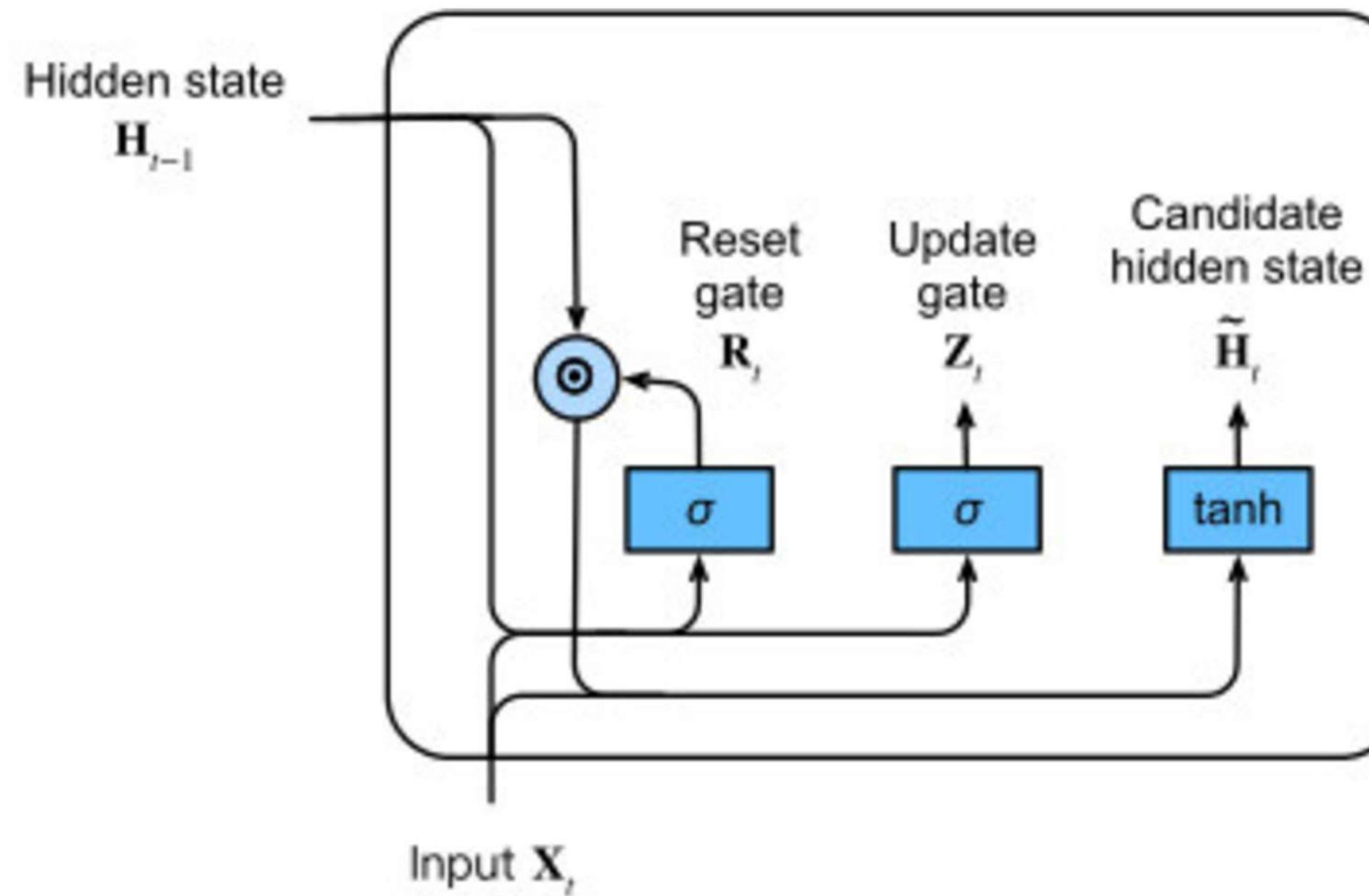
$$R_t = \sigma(W_{xr}X_t + W_{hr}H_{t-1} + b_r)$$

$$Z_t = \sigma(W_{xz}X_t + W_{hz}H_{t-1} + b_z)$$



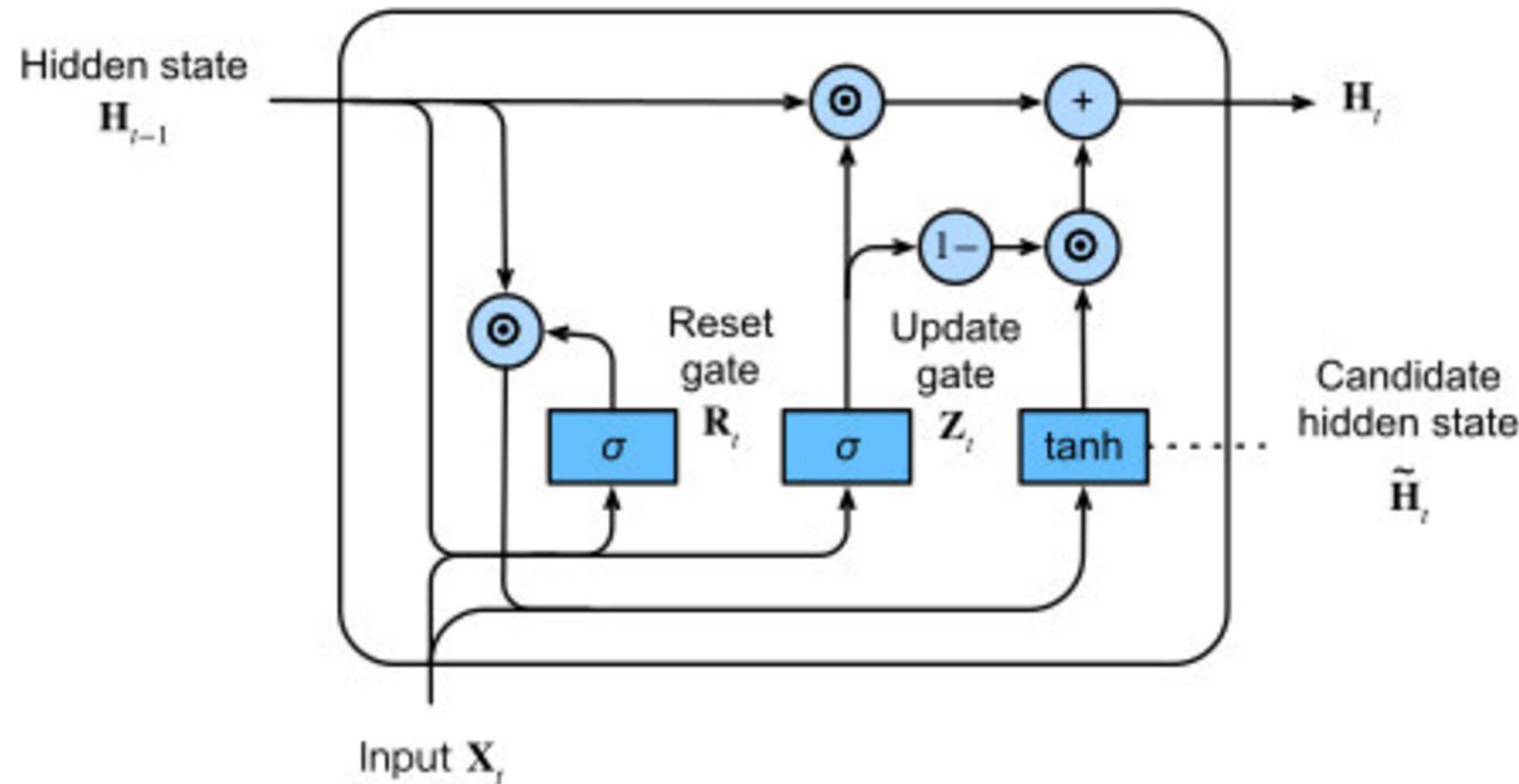
Gated Recurrent Units

$$\bar{H}_t = \tanh(W_{xh}X_t + W_{hh}(R_t \times H_{t-1}) + b_h)$$

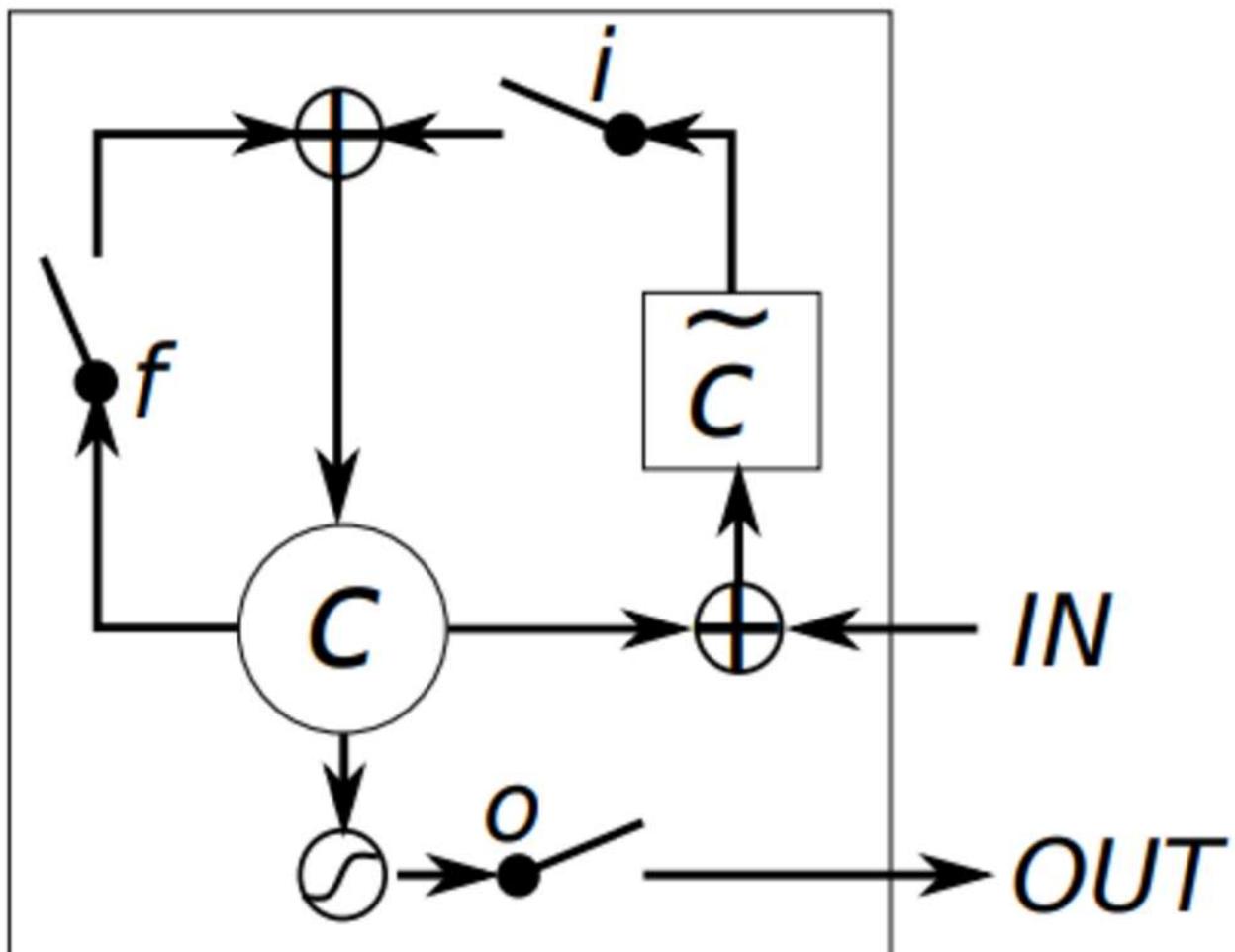


Gated Recurrent Units

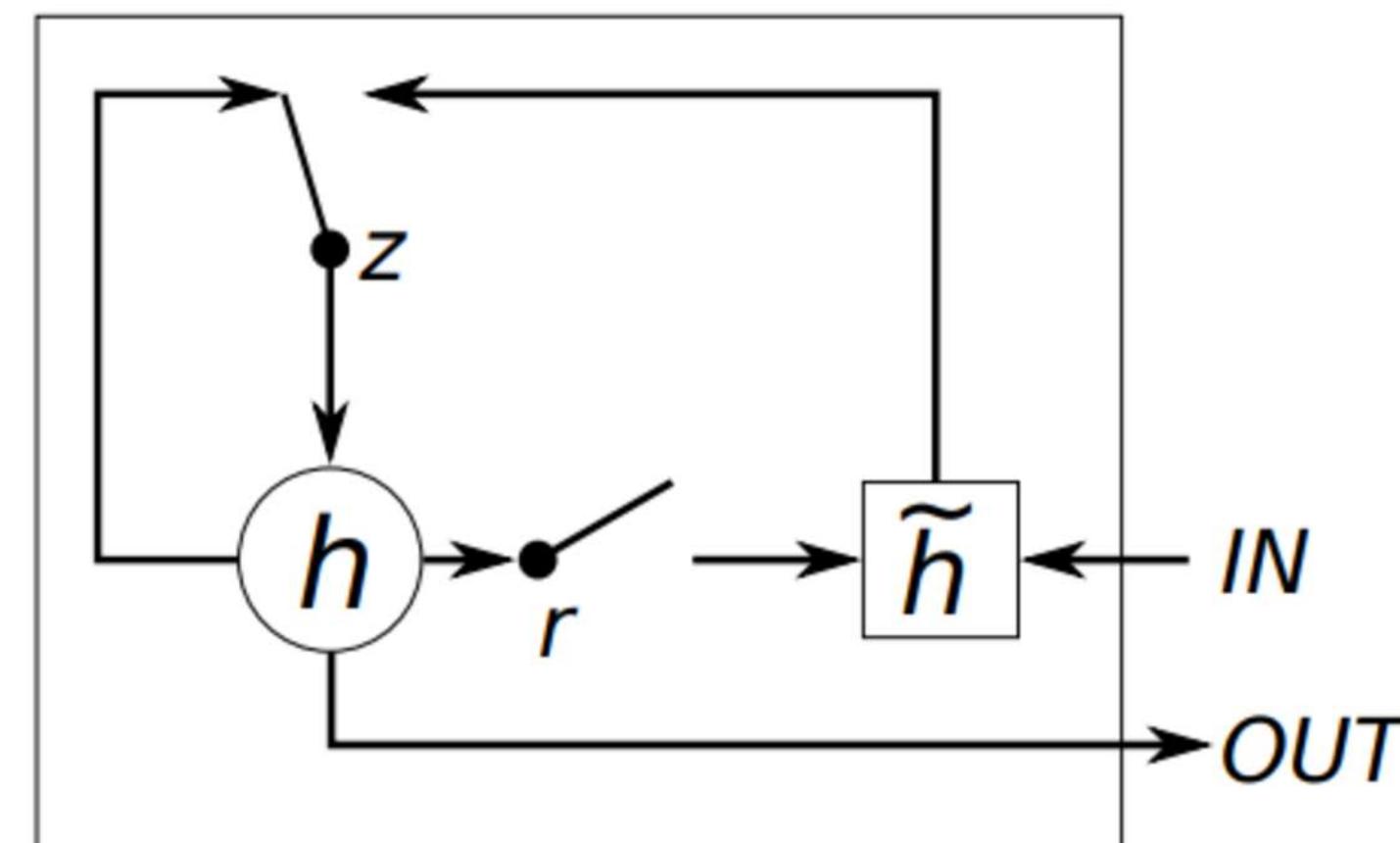
$$H_t = Z_t \times H_{t-1} + (1 - Z_t) \times \bar{H}_t$$



Gated Recurrent Units



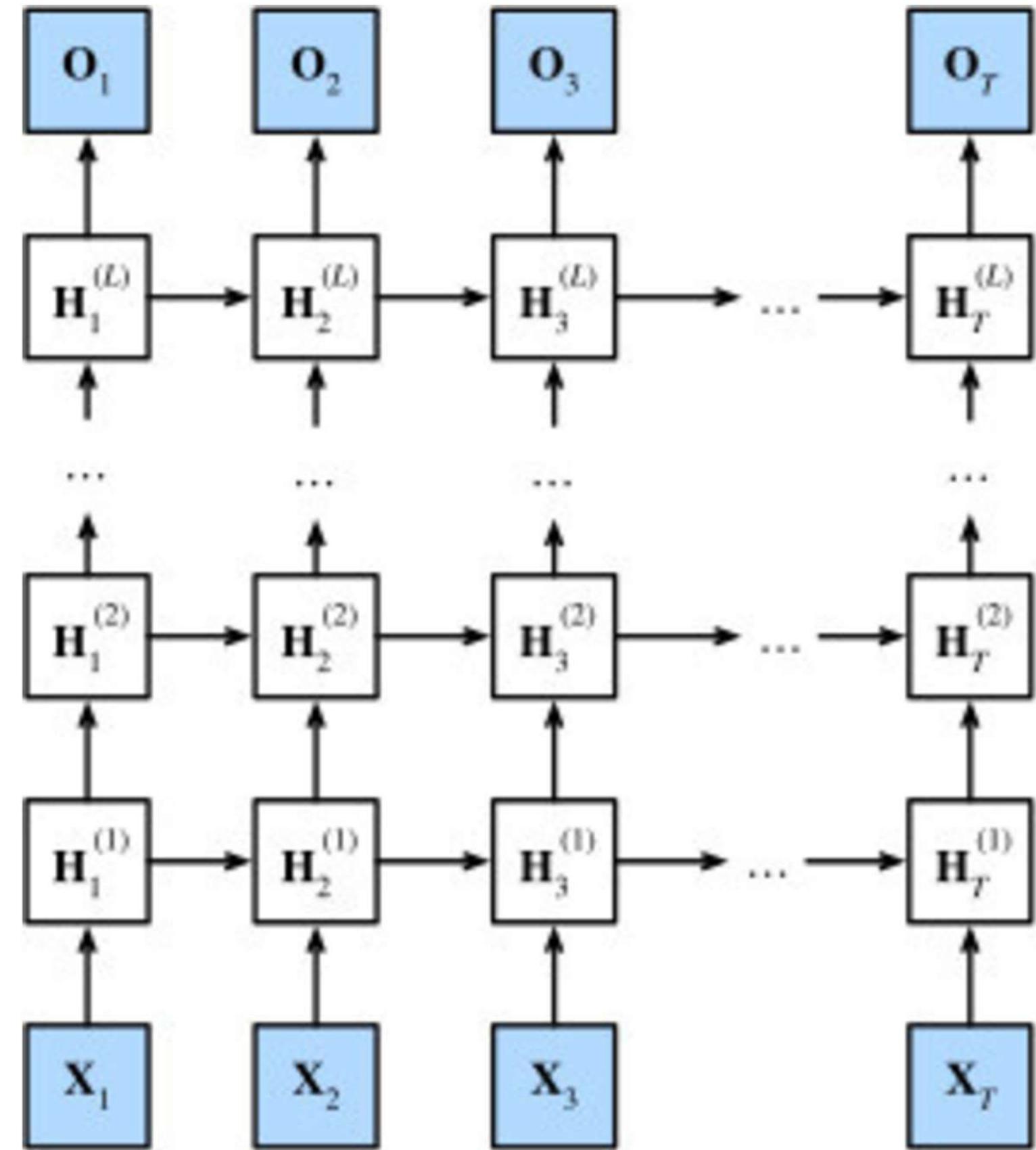
(a) Long Short-Term Memory



(b) Gated Recurrent Unit

Deep Recurrent Neural Networks

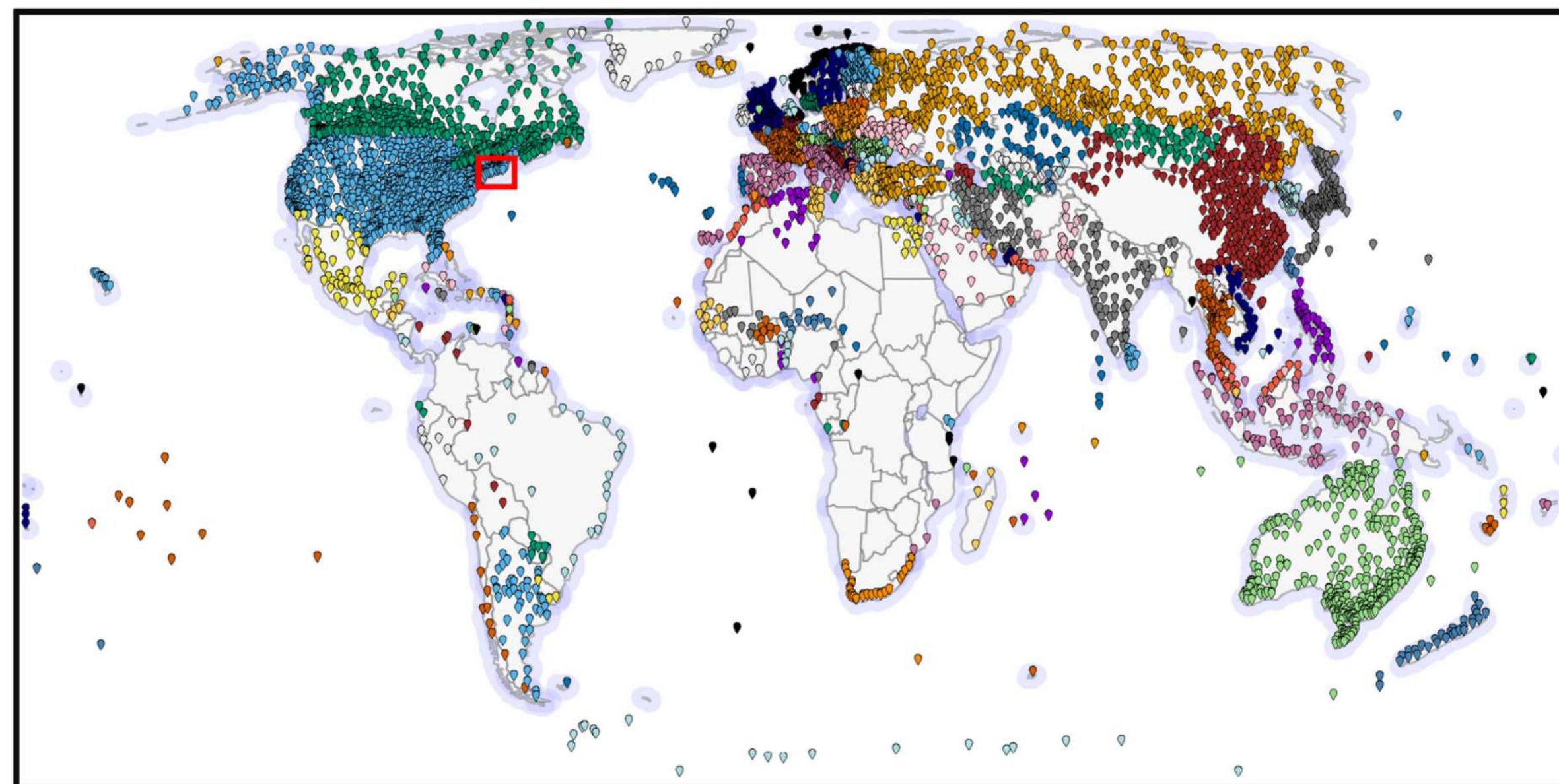
- Các mô hình RNN, LSTM và GRU gốc tập trung vào một mạng chỉ gồm input layer, hidden layer và output layer
- Một cách đơn giản để xây dựng DeepRNN là xếp chồng nhiều tầng RNN lên nhau
- Một nhược điểm lớn của DeepRNN là chi phí tính toán cao hơn và tốn nhiều bộ nhớ hơn



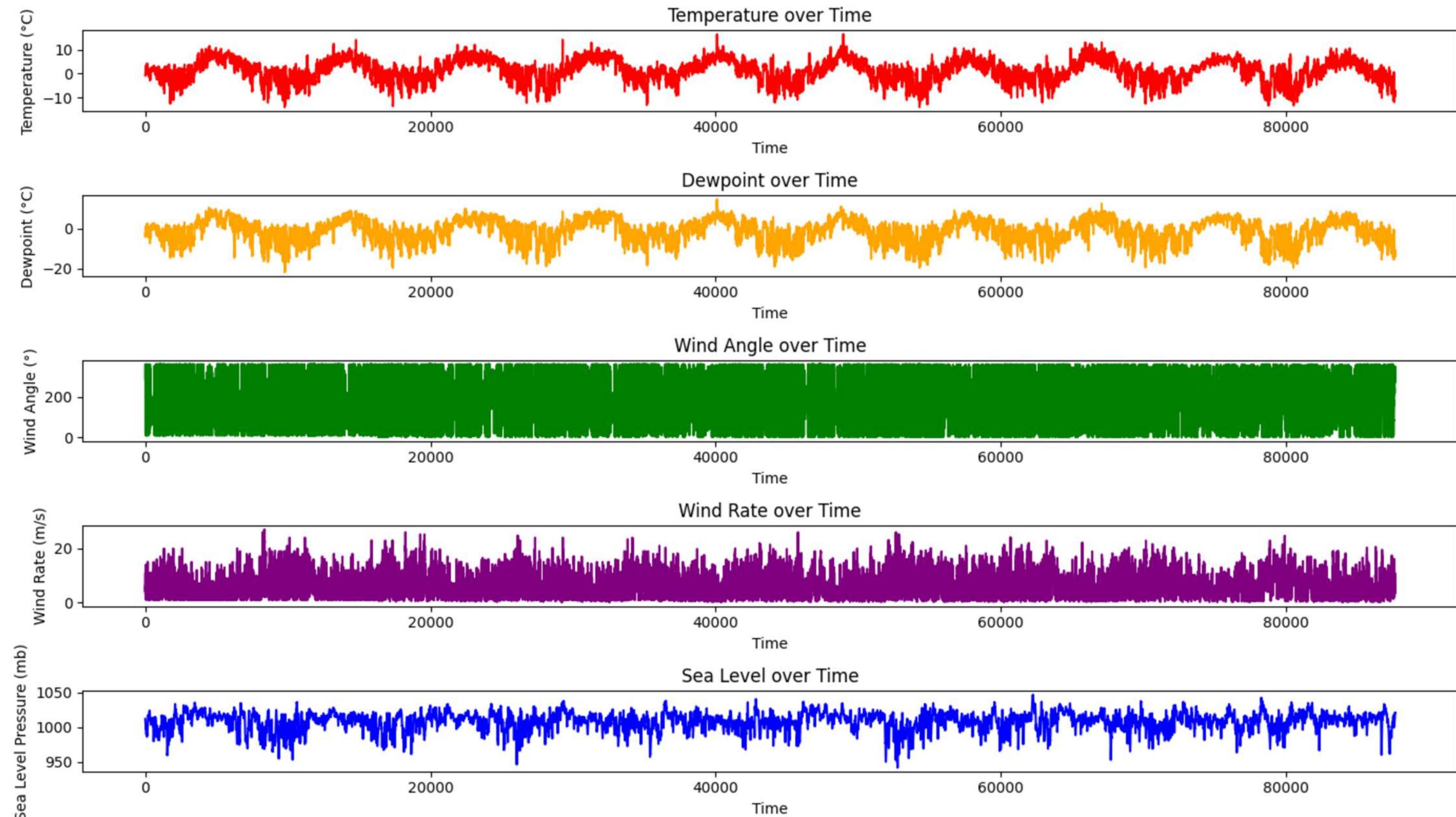
- Năm 2017, Transformer ra đời và vượt mặt RNN trong hầu hết các bài toán NLP, thay thế RNN trong việc xử lý chuỗi dài
- RNN hiện tại ứng dụng trong lĩnh vực xử lý tín hiệu thời gian thực (cảm biến, chứng khoán, sóng não,...) và là thành phần phụ/hybrid trong các kiến trúc mới

Ứng dụng trong bài toán thực tế

- Bộ dữ liệu: Weather5k
- Gồm dữ liệu thời tiết trong 10 năm của 5672 trạm khí tượng, dữ liệu thu thập cách nhau 1 tiếng



Ứng dụng trong bài toán thực tế



- Thông tin huấn luyện:

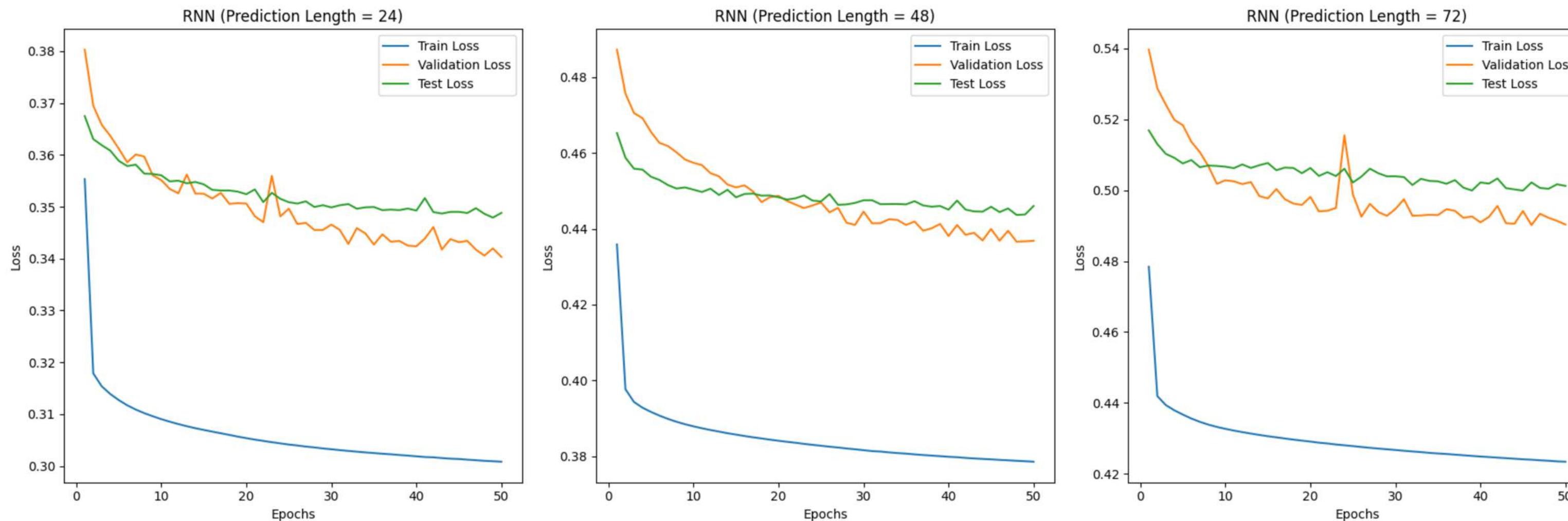
- Dữ liệu đầu vào, đầu ra: $X \in R^{1000 \times 48 \times 5}, Y \in R^{1000 \times T \times 5}$
- Prediction Length: $T \in \{24, 48, 72\}$
- Thực hiện chuẩn hóa theo từng biến
- Batch size: 1024
- Optimizer: Adam
- Learning rate: 0,0001
- Loss: MSE
- Number of Epochs: 50



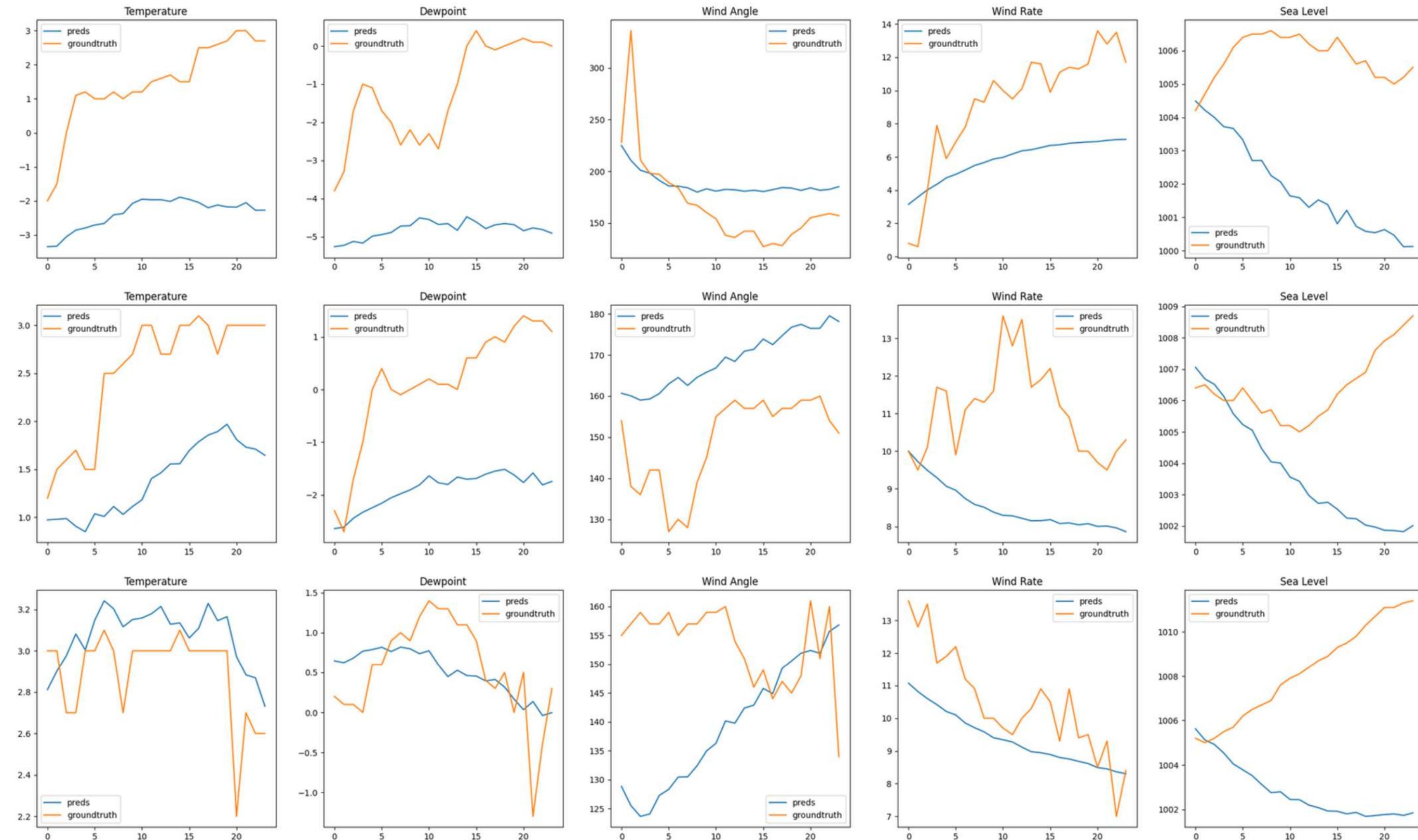
Ứng dụng trong bài toán thực tế

- RNN:

- $T = 24$ - Validation Loss: 0.3403 - Test Loss: 0.3481
- $T = 48$ - Validation Loss: 0.4366 - Test Loss: 0.4509
- $T = 72$ - Validation Loss: 0.4901 - Test Loss: 0.5055



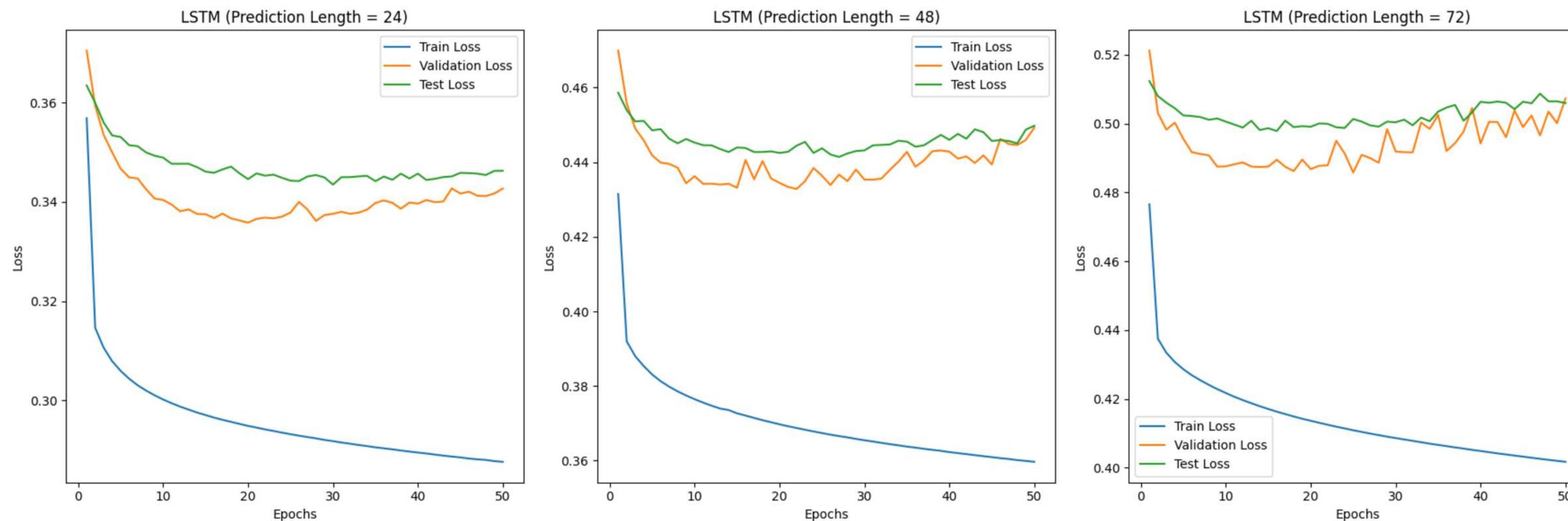
Ứng dụng trong bài toán thực tế



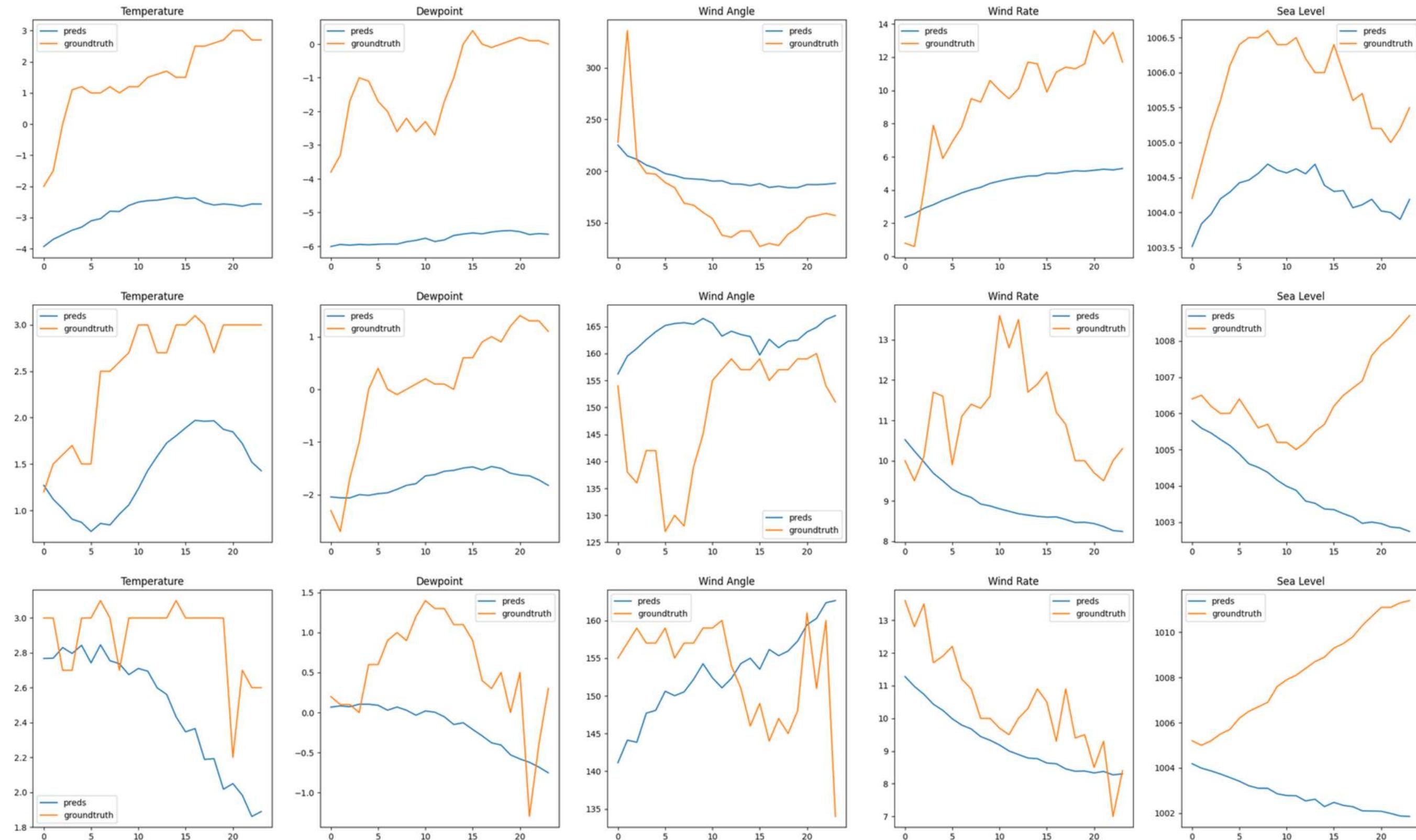
Ứng dụng trong bài toán thực tế

- LSTM:

- $T = 24$ - Validation Loss: 0.3361 - Test Loss: 0.3503
- $T = 48$ - Validation Loss: 0.4331 - Test Loss: 0.4415
- $T = 72$ - Validation Loss: 0.4873 - Test Loss: 0.4960



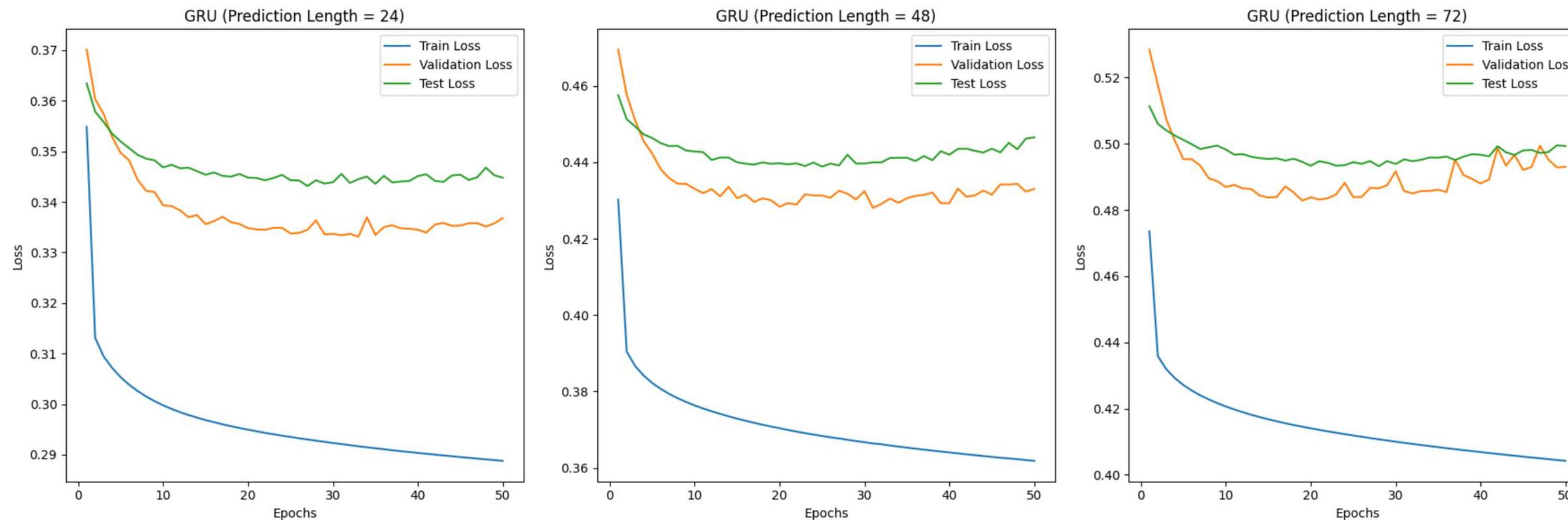
Ứng dụng trong bài toán thực tế



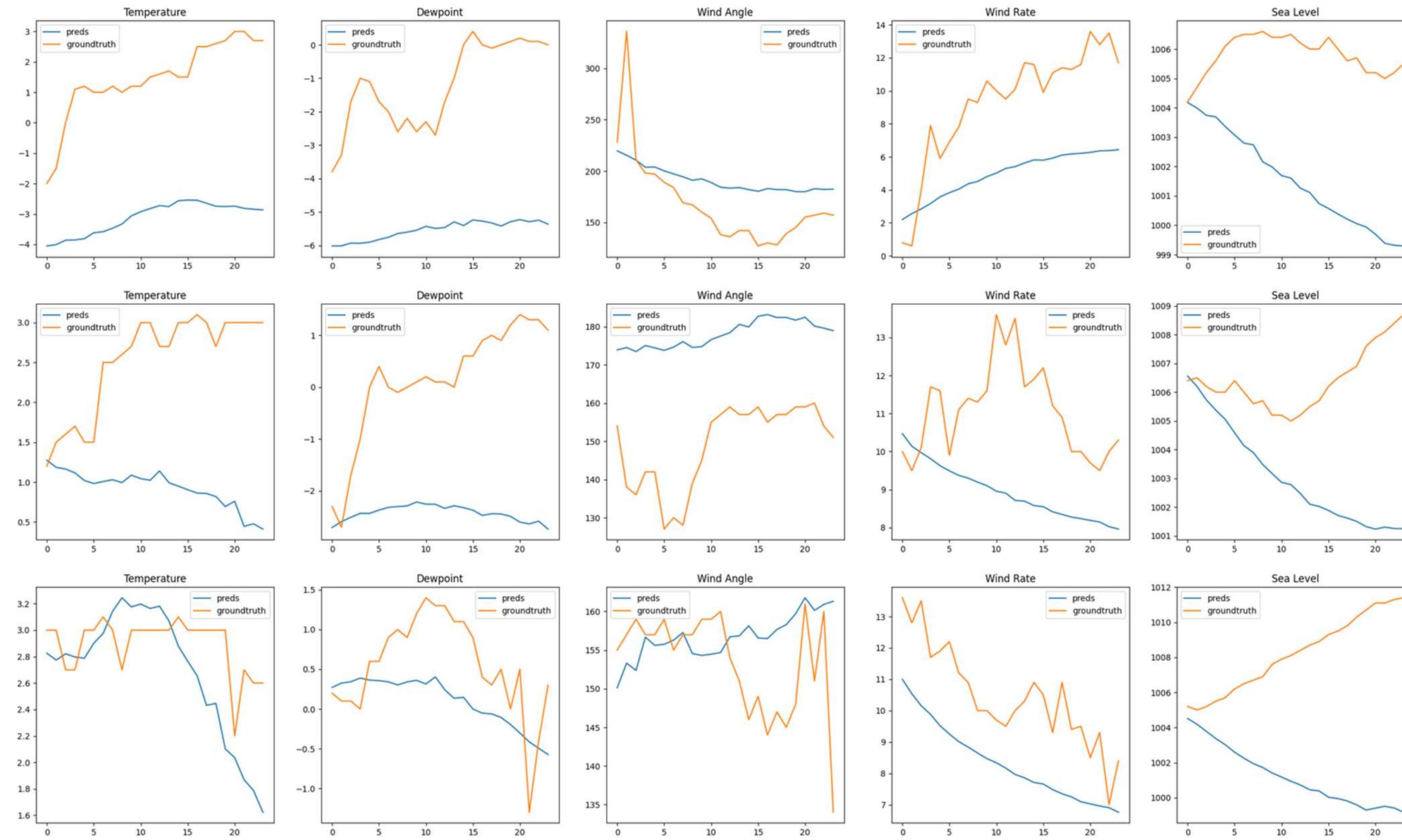
Ứng dụng trong bài toán thực tế

- GRU:

- $T = 24$ - Validation Loss: 0.3331 - Test Loss: 0.3482
- $T = 48$ - Validation Loss: 0.4283 - Test Loss: 0.4381
- $T = 72$ - Validation Loss: 0.4828 - Test Loss: 0.4934



Ứng dụng trong bài toán thực tế



Ứng dụng trong bài toán thực tế

Model	Prediction Length	Validation Loss (MSE)	Test Loss (MSE)
RNN	24	0.3403	0.3481
	48	0.4366	0.4509
	72	0.4901	0.5055
LSTM	24	0.3361	0.3503
	48	0.4331	0.4415
	72	0.4873	0.4960
GRU	24	0.3331	0.3482
	48	0.4283	0.4381
	72	0.4828	0.4934

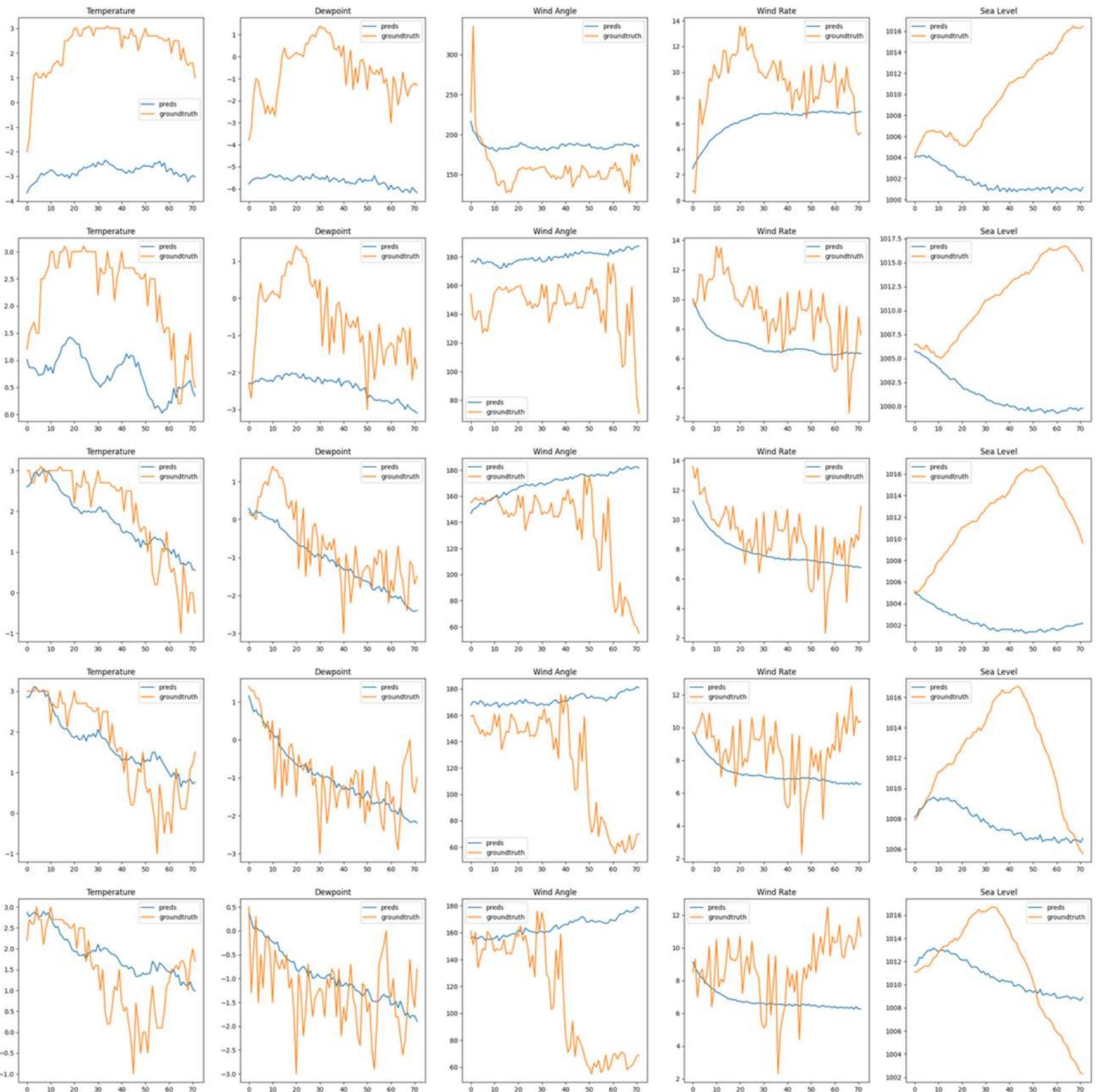


Ứng dụng trong bài toán thực tế

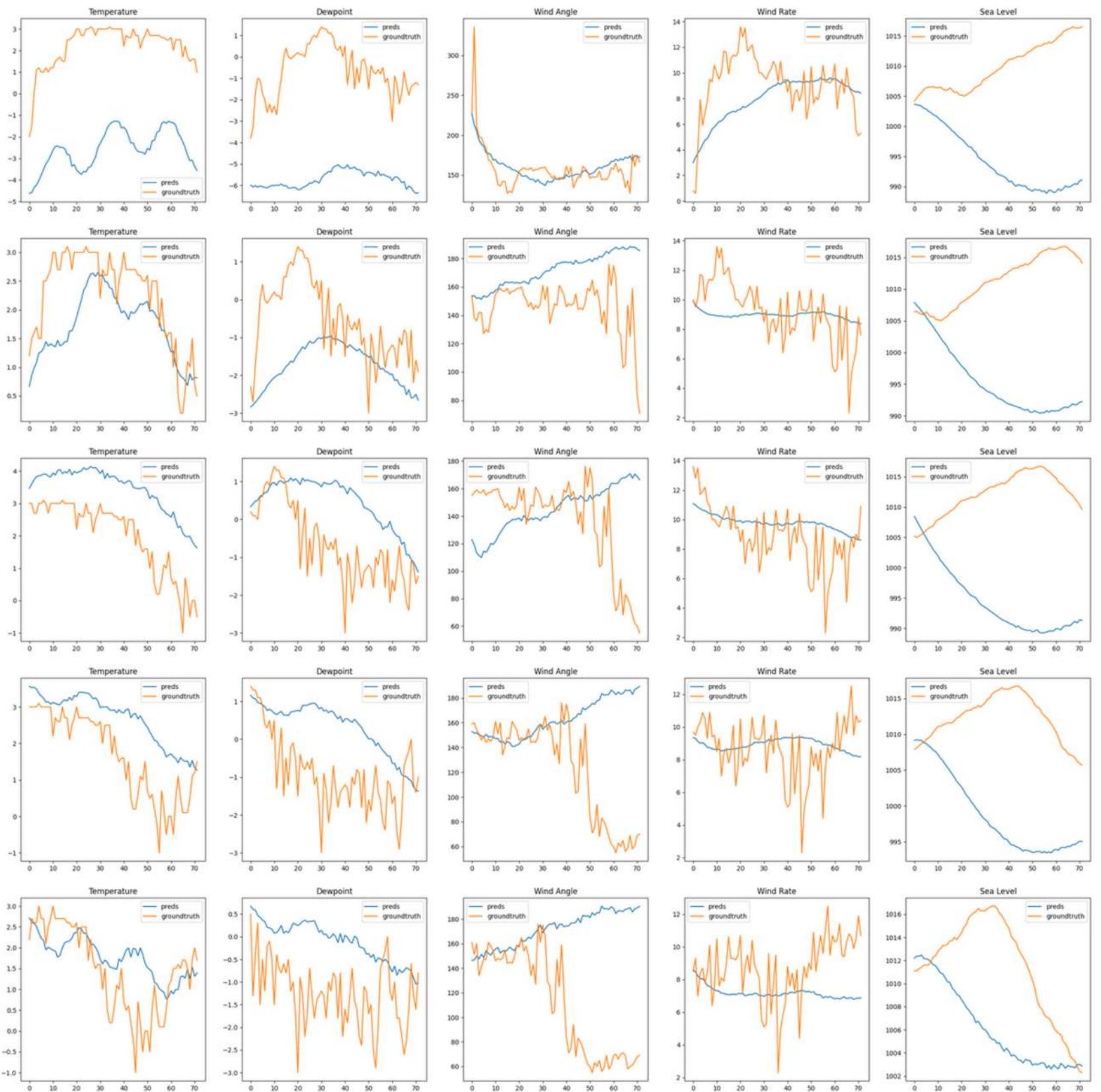
Model	Prediction Length	Test Loss (MSE)				
		Temperature	Dewpoint	Wind Rate	Wind Angle	Sea Level
RNN	24	12.01	9.74	7.02	9930.87	20.67
	48	15.55	12.48	9.08	10719.67	26.78
	72	17.43	13.99	10.18	11056.33	37.53
LSTM	24	12.08	9.70	7.06	9993.63	20.80
	48	15.22	12.22	8.90	10496.20	26.22
	72	17.11	13.73	10.01	10848.548	36.82
GRU	24	12.14	9.64	7.13	9933.72	20.68
	48	15.10	12.33	8.52	10415.38	26.02
	72	17.01	13.66	9.94	10791.67	36.63



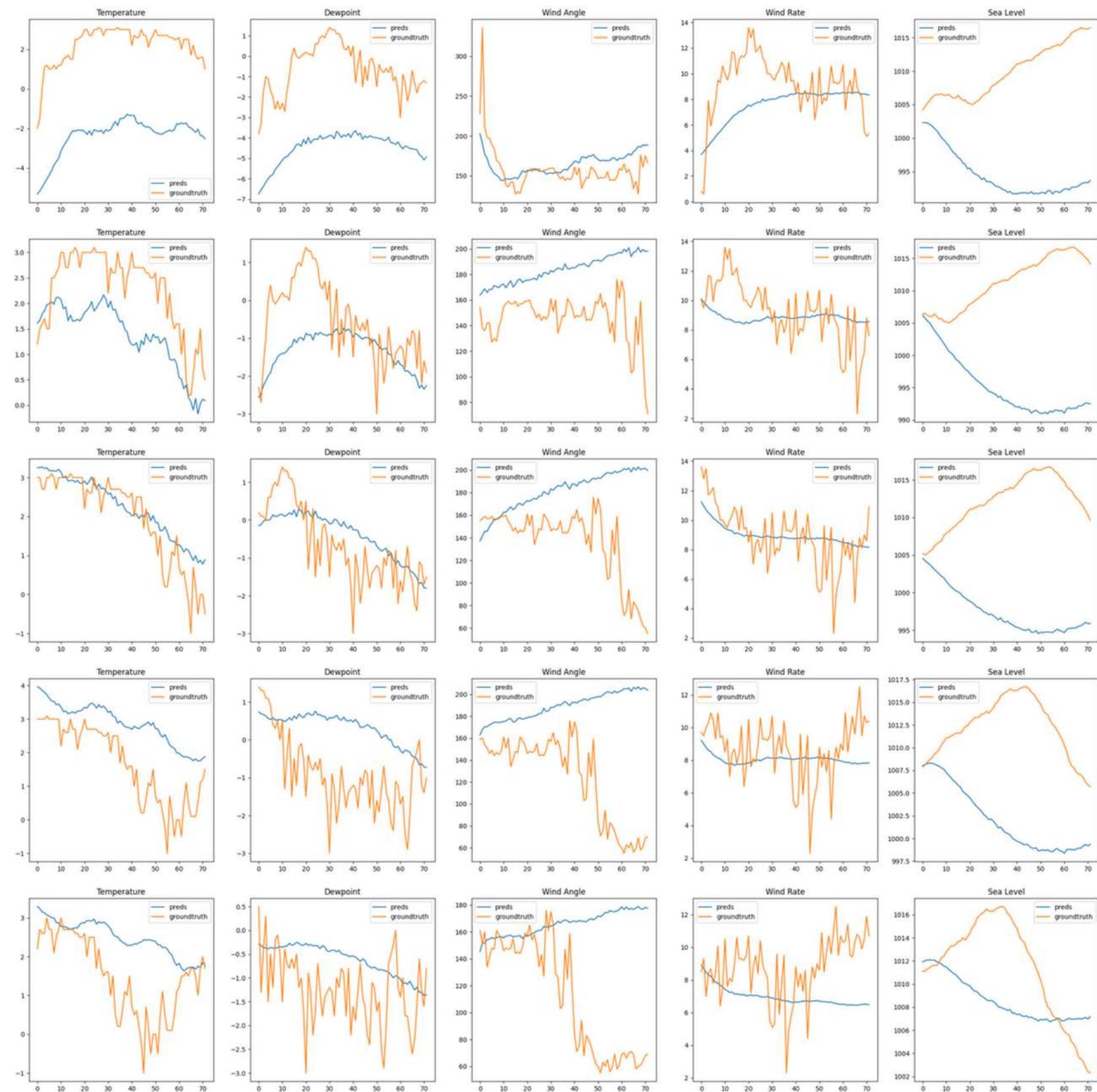
Ứng dụng trong bài toán thực tế



Ứng dụng trong bài toán thực tế



Ứng dụng trong bài toán thực tế



A large, faint watermark of the HUST logo is visible across the entire background of the slide.

HUST

THANK YOU !