

ĐẠI HỌC QUỐC GIA TP. HCM
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN



BÁO CÁO TỔNG KẾT
ĐỀ TÀI KHOA HỌC VÀ CÔNG NGHỆ SINH VIÊN NĂM 2018

Tên đề tài tiếng Việt:

**THIẾT BỊ QUAN TRẮC, DỰ BÁO THỜI TIẾT PHẠM VI NHỎ SỬ DỤNG CÔNG
NGHỆ MÁY HỌC**

Tên đề tài tiếng Anh:

LOCAL WEATHER FORECASTING SYSTEM USING NEURAL NETWORK

Khoa: Kỹ thuật Máy tính

Thời gian thực hiện: 06 tháng

Cán bộ hướng dẫn: TS. Trịnh Lê Huy

TT	Họ và tên, MSSV	Chịu trách nhiệm	Điện thoại	Email
1	Nguyễn Mạnh Thảo	Chủ nhiệm	0987612206	14520853@gm.uit.edu.vn
2	Quách Thế Hào	Thành viên	0986467627	14521124@gm.uit.edu.vn

Thành phố Hồ Chí Minh – Tháng 12/2018



ĐẠI HỌC QUỐC GIA TP. HCM
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

Ngày nhận hồ sơ	
Mã số đề tài	
(Do CQ quản lý ghi)	

BÁO CÁO TỔNG KẾT

Tên đề tài tiếng Việt:

**THIẾT BỊ QUAN TRẮC, DỰ BÁO THỜI TIẾT PHẠM VI NHỎ SỬ DỤNG CÔNG
NGHỆ MÁY HỌC**

Tên đề tài tiếng Anh:

LOCAL WEATHER FORECASTING SYSTEM USING NEURAL NETWORK

Ngày 17 tháng 12 năm 2018.
Cán bộ hướng dẫn
(Họ tên và chữ ký)

TS. Trịnh Lê Huy

Ngày 17 tháng 12 năm 2018
Sinh viên chủ nhiệm đề tài

Nguyễn Mạnh Thảo

THÔNG TIN KẾT QUẢ NGHIÊN CỨU

1. Thông tin chung:

- Tên đề tài:

Thiết bị quan trắc, dự báo thời tiết phạm vi nhỏ sử dụng công nghệ máy học

- **Chủ nhiệm:** Nguyễn Mạnh Thảo

- **Cơ quan chủ trì:** Khoa Kỹ thuật Máy tính, Trường Đại học Công nghệ Thông tin.

- **Thời gian thực hiện:** 06 tháng

2. Mục tiêu

2.1. Lý do chọn đề tài

Đề tài nghiên cứu, phát triển hệ thống thiết bị quan trắc, dự báo thời tiết trong phạm vi nhỏ, sử dụng làm nền tảng phát triển các ứng dụng liên quan đến thời tiết. Thiết bị dựa trên công nghệ máy học để đưa ra dự đoán, giúp cải thiện độ chính xác của phép dự đoán đối với từng trường hợp riêng biệt theo thời gian sử dụng, thích nghi với từng khu vực khác nhau.

Thống kê trong năm 2017, cơ cấu nền kinh tế nước ta như sau: khu vực nông, lâm nghiệp và thủy sản chiếm tỷ trọng 15,34%; khu vực công nghiệp và xây dựng chiếm 33,34%; khu vực dịch vụ chiếm 41,32%. Trong đó, các ngành nông, lâm nghiệp và thủy sản, xây dựng đóng vai trò quan trọng, chiếm tỉ trọng lớn và rất phổ biến. Đa số các mô hình trong các ngành này đều bị ảnh hưởng do thời tiết và việc dự đoán trước thời tiết một cách cục bộ mang lại lợi ích không nhỏ cũng như phòng tránh được các thiệt hại không đáng có. Cùng với sự phát triển của các mô hình Internet of Things trong các ngành, nhu cầu dự đoán thời tiết ngày càng tăng cao. Tuy nhiên, hiện tại vẫn chưa có một hệ thống đơn giản cụ thể nào đáp ứng cho nhu cầu này.

Giải pháp hiện tại để đáp ứng cho các nhu cầu hiện tại là sử dụng kênh dự báo thời tiết của quốc gia hay quốc tế để dự đoán trước thời tiết. Ưu điểm của giải pháp này là không cần phần cứng lắp thêm, không tốn chi phí do thông tin được công bố trên các kênh truyền thông miễn phí. Tuy nhiên, giải pháp này có khuyết điểm là không hoàn toàn chính xác cho một khu vực cục bộ cụ thể.

Sau khi tiến hành tìm hiểu, nhóm nghiên cứu quyết định thực hiện đề tài Thiết bị dự báo thời tiết trong phạm vi nhỏ sử dụng công nghệ Neural Network được mục đích quan trắc và dự báo thời tiết trong một khu vực nhỏ, cá nhân hóa người dùng, phục vụ cho các hộ gia đình và doanh nghiệp nhỏ, điều mà các hệ thống quan trắc lớn, quy mô quốc gia, khó có thể làm được.

2.2. Mục tiêu tổng quan

Xây dựng thành công hệ thống có khả năng tự nó thu thập các giá trị môi trường và đưa ra dự báo mà không phụ thuộc vào các thông tin dữ liệu từ bên thứ ba. Có thể hoạt động trong thời gian dài ở ngoài môi trường thực tế, đóng gói sản phẩm để thiết bị có thể hoạt động được dưới trời mưa bão.

Đề tài hướng đến sử dụng Machine Learning, cụ thể là mô hình Back-propagation Neural Network giải quyết bài toán dự đoán mưa phạm vi nhỏ. Sử dụng ưu điểm của mô hình Back-propagation Neural Network cải thiện độ chính xác của dự đoán theo thời gian sử dụng hệ thống.

Sử dụng công nghệ truyền dữ liệu LoRa để áp dụng vào việc truyền nhận tín hiệu, giúp gia tăng khoảng cách lắp đặt hệ thống, theo dõi được tình hình tại khu vực có khoảng cách xa nơi lắp đặt server trung tâm. Không những tăng khoảng cách truyền nhận dữ liệu, công nghệ LoRa còn giúp tiết kiệm năng lượng sử dụng rất phù hợp với hệ thống. Bên cạnh đó, nhờ kết hợp với pin mặt trời, thời gian sử dụng của thiết bị quan trắc của hệ thống được kéo dài, qua đó giúp hệ thống có thể lắp đặt được ở nhiều khu vực khác nhau, tăng khả năng mở rộng hệ thống quan trắc.

Bên cạnh đó nhóm cũng đưa hệ thống lên môi trường Internet bằng cách thiết kế Gateway có kết nối Wifi, đưa dữ liệu lên VPS để xử lý với thuật toán Machine Learning. Ngoài ra, VPS này cũng chạy một Web Server, cung cấp cho người một trang Web quản lý với giao diện thân thiện.

Trên tất cả, vấn đề lớn nhất mà nhóm đang tập trung giải quyết chính là độ chính xác và độ tin cậy của hệ thống phải được tối ưu nhất có thể và phải đặt lên hàng đầu.

2.3. Mục tiêu cụ thể

Đối tượng nghiên cứu của đề tài như sau:

- Hệ thống thu thập thông số môi trường SKU:SEN0186
- Pin năng lượng mặt trời mono, pin li-ion và mạch sạc
- Cảm nhiệt độ, độ ẩm SHT10, cảm biến áp suất BMP180
- Cơ sở lý luận về dự báo thời tiết và dữ liệu
- Mô hình Machine Learning Back-propagation Neural Network
- Module truyền nhận LoRa RFM96
- Module Wifi ESP8266
- Giao thức MQTT
- Web Server với ngôn ngữ NodeJS
- Hệ cơ sở dữ liệu MongoDB

Với mục tiêu xây dựng được một hệ thống phần cứng thu thập dữ liệu môi trường một cách độc lập, được đóng gói để hoạt động trong môi trường mưa gió khắc nghiệt. Bao gồm:

- Trạm thu dữ liệu: Thu thập các giá trị như nhiệt độ, độ ẩm, áp suất khí quyển, hướng gió, sức gió và lượng mưa. Sử dụng pin năng lượng mặt trời để duy trì hoạt động.
- Gateway: Thu thập dữ liệu từ trạm thu dữ liệu và đưa lên VPS thông qua Wifi. Nhận dữ liệu dự đoán thời tiết từ VPS và hiển thị lên LCD.

Nghiên cứu về cơ sở lý thuyết dự báo thời tiết, tìm hiểu các bài báo về thời tiết. Kết hợp với việc xử lý, phân tích giá trị thực tế thu thập được trong thời gian dài. Từ đó đưa ra cơ sở lý luận tin cậy cho việc dự báo thời tiết.

Nghiên cứu và xây dựng ứng dụng dự đoán thời tiết cục bộ sử dụng mô hình Back-propagation Neural Network. So sánh với hệ thống sử dụng Fuzzy Logic trước đây. Tăng tỉ lệ chính xác lên 60%.

Xây dựng hệ thống quản lý dữ liệu với MongoDB để quản lý dữ liệu cảm biến, kết quả dự đoán và dữ liệu người dùng.

Xây dựng một Web Server trên VPS Linux với giao diện thân thiện và hoạt động ổn định.

Cụ thể:

- Dữ liệu thu thập môi trường ít nhất trong khoảng 3 tháng, xử lý dữ liệu và để mô hình Back-propagation Neural Network học tập dữ liệu ban đầu này.
- Xây dựng ứng dụng dự đoán thời tiết cục bộ sử dụng mô hình Back-propagation Neural Network để dự đoán thời tiết. Ứng dụng sử dụng 7 biến đầu vào bao gồm: nhiệt độ, độ ẩm, áp suất, giờ trong ngày, tháng trong năm, hướng gió và sức gió để dự đoán và thêm một biến đầu vào là lượng mưa để học.
- Giao tiếp với các cảm biến. Thực hiện đọc dữ liệu từ các cảm biến.
- Xây dựng nguồn nuôi cho hệ thống thu dữ liệu bằng pin li-ion, dùng pin năng lượng mặt trời để duy trì hoạt động của pin.
- Thực hiện giao tiếp giữa trạm thu dữ liệu và board trung tâm bằng sóng LoRa.
- Xây dựng một trang Web bằng ngôn ngữ HTML, Css và NodeJS, trang Web có giao diện đẹp, tính tương tác người dùng cao.
- Tạo và đăng kí được một tên miền trên No-IP.
- Xây dựng ứng dụng quản lý, truy xuất trên hệ cơ sở dữ liệu MongoDB.

3. Tính mới và sáng tạo

3.1. Phân tích hiện trạng

3.1.1. Tình hình nghiên cứu trong nước

Việt Nam là một nước nằm trong khu vực có thời tiết thay đổi phức tạp, biến hóa khôn lường, thường xuyên xảy ra thiên tai lũ lụt, đặc biệt là các tỉnh thành miền Trung và miền Bắc nước ta, nên việc xây dựng các trạm quan trắc dự báo thời tiết đang rất được chú trọng ở Việt Nam.

Các trạm quan trắc thời tiết ở Việt Nam, hầu như có qui mô lớn trên diện rộng, chưa áp dụng phát triển ở từng địa phương khu vực. Các trang thiết bị đều được

nhập từ nước ngoài với chi phí đắt đỏ và độ chính xác không cao do phạm vi quan trắc còn lớn và số lượng trạm quan trắc chưa nhiều

Trong cuộc thi Nhà sáng tạo Việt Nam với Intel Galileo năm 2015 (VMIG 2015), nhóm sinh viên Trường Đại học Khoa học Tự nhiên đã giành giải Ba với đề tài "Trạm quan trắc môi trường và cảnh báo sớm thiên tai". Hệ thống gồm 3 thành phần chính: hệ thống cảm biến xử lý dữ liệu, board xử lý trung tâm là Intel Galileo và cuối cùng là một Web hiển thị thông tin và đưa ra cảnh báo khi vượt mức. Đề tài nghiên cứu của nhóm bạn có ứng dụng rất cao và được hội đồng đánh giá cao, nhưng hệ thống vẫn còn một số bất cập đó là, hệ thống chỉ mang tính cảnh báo tại vị trí và thời điểm hiện tại chưa có tính dự đoán thời tiết xảy ra tương lai.

Công ty Mimosatek - Giải Nhất Nhóm Khởi Nghiệp Cuộc Thi Venture Cup 2015 với sản phẩm "Hệ thống quản lý tưới chính xác", một thiết bị dùng để đo các giá trị thời tiết, tính toán phân tích thông tin để biết được lượng nước cần tưới cho từng loại cây trồng, tất cả những việc này được quản lý và điều khiển thông qua điện thoại thông minh. Hệ thống này chỉ đo giá trị thời tiết tại thời điểm hiện tại, để thực hiện tưới nước hợp lý và cảnh báo khi môi trường bất lợi chưa có mang tính dự báo thời tiết tiếp sắp tới.

3.1.2. Tình hình nghiên cứu quốc tế

Tác giả Saktaya Suksri and Warangkhan Kimpan, bài báo "Neural Network training model for weather forecasting using Fireworks Algorithm". Công bố tại International Computer Science and Engineering Conference (ICSEC) ngày 14-17 Dec. 2016

Bài báo trình bày phương pháp sử dụng mô hình Artificial Neural Networks with supervised learning paradigm để dự đoán giá trị nhiệt độ trung bình trong ngày với dữ liệu đầu vào lấy từ trạm khí tượng thủy văn đặt ở Bangkok. Mô hình Artificial Neural Network được train bởi thuật toán Fireworks - thuật toán tối ưu hóa việc train ANN được phát triển bởi Swarm Intelligence Algorithm. Tuy nhiên, bài báo chỉ tập trung phát triển thuật toán và nghiên cứu chứ không phát triển thành sản phẩm.

Trường đại học, bằng Master. Tác giả Andrew Culclasure, đề tài "Using Neural Networks to Provide Local Weather Forecasts", Công bố tại Electronic Theses & Dissertations Georgia Southern University Spring 2013.

Đề tài tập trung vào việc thực hiện khảo sát trên các mô hình dự đoán thời tiết sử dụng ANN đã có tại thời điểm thực hiện. Ngoài ra, đề tài cũng giới thiệu một thí nghiệm của tác giả về việc sử dụng ANN để dự đoán nhiệt độ.

Tác giả Kumar Abhishek, bài báo "Weather Forecasting Model using Artificial Neural Network", được đăng trên tạp chí Procedia Technology Volume 4, 2012, Pages 311-318.

Bài báo xem xét độ khả thi của phương pháp sử dụng ANN với các thông số hàm truyền, số lượng lớp ẩn và neural khác nhau trong việc dự đoán thời tiết, cụ thể hơn là dự đoán nhiệt độ tối đa trong một năm.

3.2. Phân tích các công nghệ

Công nghệ máy học

Máy học hay còn gọi là học máy là tên được đặt ra vào năm 1959 bởi Arthur Samuel. Phát triển từ nghiên cứu nhận dạng mẫu và lý thuyết học tính toán trong trí tuệ nhân tạo, máy học khám phá nghiên cứu và xây dựng các thuật toán có thể học và đưa ra dự đoán về dữ liệu - các thuật toán đó vượt qua các hướng dẫn chương trình tĩnh nghiêm ngặt bằng cách đưa ra các dự đoán hoặc quyết định dựa trên dữ liệu, thông qua việc xây dựng một mô hình từ các đầu vào mẫu. Máy học được sử dụng trong các công việc điện toán trong đó việc thiết kế và lập trình các thuật toán rõ ràng với hiệu suất tốt với con người thì rất khó khăn hoặc không khả thi; các ứng dụng ví dụ bao gồm lọc email, phát hiện kẻ xâm nhập mạng làm việc theo hướng vi phạm dữ liệu, nhận dạng ký tự quang học (OCR), học cách xếp hạng và thị giác máy tính.

Máy học có liên quan chặt chẽ với (và thường trùng lặp với) thống kê tính toán, cũng tập trung vào việc dự đoán thông qua việc sử dụng máy tính. Nó có mối quan hệ chặt chẽ với tối ưu hóa toán học, cung cấp các phương thức, lý thuyết và lĩnh vực ứng dụng cho lĩnh vực này. Máy học đôi khi được kết hợp với khai thác dữ liệu, trong đó trường con sau tập trung nhiều hơn vào phân tích dữ liệu khám phá và được gọi là học tập không giám sát. máy học cũng có thể không được giám sát và được sử dụng để tìm hiểu và thiết lập hồ sơ hành vi cơ bản cho các thực thể khác nhau và sau đó được sử dụng để tìm ra sự bất thường có ý nghĩa.

Trong lĩnh vực phân tích dữ liệu, máy học là một phương pháp được sử dụng để đưa ra các mô hình và thuật toán phức tạp cho vay để dự đoán; trong sử dụng thương mại, điều này được gọi là phân tích dự đoán. Những mô hình phân tích này cho phép các nhà nghiên cứu, nhà khoa học dữ liệu, kỹ sư và nhà phân tích "đưa ra các quyết định và kết quả đáng tin cậy, có thể lặp lại" và khám phá "những hiểu biết ẩn giấu" thông qua việc học hỏi từ các mối quan hệ và xu hướng lịch sử trong dữ liệu.

Việc ứng dụng máy học trong lĩnh vực dự đoán thời tiết vẫn còn khá mới. Ngoài máy học, logic mờ hay Fuzzy logic cũng được ứng dụng trong lĩnh vực dự đoán thời tiết, tuy nhiên độ chính xác của dự đoán và khả năng thích ứng với môi trường của máy học vượt trội hơn hẳn.

Công nghệ giao tiếp truyền thông LoRa

LoRa, viết tắt của Long Range Radio, là một công nghệ truyền thông không dây được phát triển bởi Cycleo và được Semtech mua vào năm 2012. LoRa sử dụng một kỹ thuật điều chế có tên là Chirp Spread Spectrum. Theo công bố của SemTech, kỹ thuật này làm giảm yêu cầu về độ phức tạp và độ chính xác của các thiết bị thu cần đạt được để giải mã dữ liệu. Ngoài ra, LoRa không yêu cầu công suất truyền cao để truyền đến thiết bị thu ở khoảng cách xa khiến thiết bị thu có thể nhận dữ liệu ngay cả khi cường độ tín hiệu nhận được thấp hơn nhiều của môi trường xung quanh.

Công nghệ bao gồm hai phần - Lora, lớp vật lý và LoRaWan, lớp giao thức. LoRaWan là một giao thức dựa trên LoRa được phát triển bởi LoRa Alliance. Nó được sử dụng trong các băng tần radio công nghiệp, khoa học và y tế (ISM). Cấu trúc mạng LoRaWan thường là mô hình sao. Gateway là cầu nối chuyển tiếp các gói tin giữa các thiết bị đầu cuối và máy chủ. Các gateway kết nối với internet thông qua các kết nối IP thông thường. Trong khi đó, các thiết bị đầu cuối không dây giao tiếp với một hoặc nhiều gateway.

Một gateway LoRaWan duy nhất có thể bao phủ khu vực có bán kính lên đến 10Km. Với những lợi ích từ khoảng cách của LoRaWan, Mạng lưới Things bao phủ Amsterdam chỉ với 10 cổng và có giá \$1200.

Ưu điểm vượt trội của LoRa khi so sánh với các công nghệ truyền dữ liệu không dây khác đó là khoảng cách giao tiếp và tiết kiệm năng lượng. Những đặc điểm này rất thích hợp ứng dụng trong đề tài, nơi mà các thiết bị quan trắc được đặt ngoài trời, xa gateway và sử dụng pin năng lượng mặt trời. Bảng so sánh chi tiết ưu/nhược điểm của các công nghệ truyền dữ liệu không dây được thể hiện trong hình 1.

Cơ sở dữ liệu MongoDB

MongoDB là một cơ sở dữ liệu định hướng tài liệu đa nền tảng. Nó cho hiệu suất cao hơn và dễ dàng mở rộng hơn các hệ quản trị cơ sở dữ liệu quan hệ. Các tài liệu của MongoDB có cấu trúc giống như JSON. MongoDB được MongoDB Inc giới thiệu lần đầu tiên vào tháng 2 năm 2009. Sau đó, MongoDB nhanh chóng phát triển thành một trong những cơ sở dữ liệu NoQuery phổ biến nhất.

Phiên bản hiện tại của MongoDB là 4.1.1 được xuất bản tại Git: github.com/mongodb/mongo.

Một số tính năng chính của MongoDB

- Schema less
- High availability by cluster
- Structure of a single object is clear
- No complex joins
- Deep query-ability
- Tuning
- Ease of scale-out
- Uses internal memory for storing the (windowed) working set, enabling faster access of data

Các ứng dụng của MongoDB

- Big Data
- Content Management

Giao thức MQTT

MQTT được phát triển bởi IBM và Eurotech, phiên bản mới nhất là MQTT 3.1.1. MQTT (Message Queuing Telemetry Transport) là một giao thức xuất bản/đăng ký thường được sử dụng cho các thiết bị Internet of Things với băng thông thấp, độ tin cậy cao và khả năng được sử dụng trong các mạng không ổn định. Nó dựa trên một Broker và được thiết kế mở và không dành riêng cho bất kỳ ứng dụng nào, rất đơn giản và dễ tích hợp. MQTT phù hợp cho các ứng dụng M2M (Mobile to Mobile), WSN (Wireless Sensor Networks) hoặc IoT (Internet of Things). Hình 2 cho thấy cấu trúc của một hệ thống MQTT cơ bản.

Có 3 cấp độ QoS (Chất lượng dịch vụ) trong MQTT:

- QoS 0 - Gần như một lần: Broker/Client gửi dữ liệu một lần, chỉ giao thức TCP/IP xác nhận việc nhận thành công
- QoS 1 - Ít nhất một lần: Broker/Client gửi ít nhất một gói nhận thành công cho người gửi
- QoS 2 - Chính xác một lần: Broker/Client chỉ gửi một gói cho người gửi để xác nhận rằng quá trình nhận thành công

Hình 3 biểu diễn các khái niệm QoS này.

3.3. Tính mới và sáng tạo của đề tài

Nhìn chung, bài toán dự báo thời tiết phạm vi nhỏ đã và đang nhận được rất nhiều sự quan tâm cũng như đầu tư nghiên cứu cả trong và ngoài nước và cũng đã có những thành quả nhất định.

Tuy nhiên, điểm mới của đề tài là trực tiếp phát triển hệ thống thành một hệ thống dự báo thời tiết độc lập mà hạn chế sự phụ thuộc vào các thông tin từ Internet. Hệ thống tự nó là một thiết bị có thể dự báo thời tiết có khả năng di động cao, độ chính xác cao.

Nhóm ứng dụng công nghệ máy học vào hệ thống giải quyết bài toán dự báo thời tiết, đây là một hướng đi mới trong tình hình nghiên cứu về thời tiết trong và ngoài nước. Nhóm hi vọng nếu thành công thì thiết bị có thể thay thế hoặc hỗ trợ trong việc dự đoán thời tiết.

Sử dụng công nghệ sóng LoRa để truyền dẫn, đây là một công nghệ mới, truyền được dữ liệu ở khoảng cách xa (khoảng 15 Km).

Bên cạnh đó nhóm cũng đã đưa được hệ thống lên môi trường Internet of Things với việc gửi dữ liệu lên trên mạng. Kết hợp với việc giảm tối thiểu giá thành là một trong những điểm mạnh của đề tài.

4. Tóm tắt kết quả nghiên cứu:

4.1. Đặc tả kỹ thuật

Thiết bị quan trắc môi trường sử dụng pin mặt trời để cung cấp năng lượng, sử dụng pin 18650 để duy trì hoạt động vào ban đêm và những ngày không có nắng. Hệ thống

có thể duy trì hoạt động liên tục trong 48 tiếng không có nắng để tránh việc gián đoạn vào những ngày mưa bão.

Hệ thống được đóng gói để có thể hoạt động ngoài trời, chịu được thời tiết mưa bão.

Các thông số như tốc độ gió, hướng gió, lượng mưa, áp suất không khí, nhiệt độ, độ ẩm, lượng mưa tại nơi cần dự đoán được thu thập và gửi về máy chủ để xử lý mỗi 5 phút.

Mô hình máy học hoạt động ở máy chủ được định thời kích hoạt dự đoán mỗi 1 giờ 1 lần và học lại dữ liệu mới mỗi 24 giờ.

Hệ thống có giao diện web cho người dùng có thể truy cập và theo dõi kết quả dự đoán cũng như thông tin về thời tiết hiện thời tại nơi đặt thiết bị.

Kết quả dự đoán mưa trước 2 giờ đạt trên 63% độ chính xác so với kết quả thực tế thu được từ chính thiết bị quan trắc.

4.2. Thiết kế hệ thống

Tổng quan hệ thống được thiết kế gồm 4 thành phần chính: Các thiết bị quan trắc môi trường (Nodes), gateway, máy chủ (Server), giao diện người dùng (User). Liên kết với nhau như trong hình 4. Chức năng từng khối có thể được mô tả như sau:

Các thiết bị quan trắc môi trường chịu trách nhiệm thu thập dữ liệu môi trường tại nơi cần dự báo thời tiết và gửi dữ liệu này đến máy chủ thông qua các gateway sau mỗi 5 phút. Dữ liệu thu thập bao gồm hướng gió, tốc độ gió, nhiệt độ, độ ẩm, áp suất, lượng mưa. Đầu ra của bộ phận này bao gồm hướng không khí, tốc độ không khí trong 1 phút, tốc độ không khí trong 5 phút, nhiệt độ, lượng mưa trong 1 giờ, lượng mưa trong 24 giờ, độ ẩm và không khí.

Gateway có trách nhiệm chuyển tiếp các gói từ các thiết bị quan trắc môi trường đến máy chủ để xử lý qua Internet, Quản lý các các thiết bị quan trắc trong khu vực được bảo vệ LoRa, xác nhận gói và yêu cầu các các thiết bị quan trắc gửi lại gói khi nhận gói không thành công.

Máy chủ nhận các gói dữ liệu từ Gateway, xử lý và lưu dữ liệu thời tiết vào cơ sở dữ liệu. Sau đó, dữ liệu này được sử dụng để dự đoán lượng mưa trong 2 giờ trước cho từng khu vực riêng lẻ mỗi giờ một lần. Dự đoán được kích hoạt một lần mỗi giờ một lần và quá trình học tập dữ liệu mới được kích hoạt một lần mỗi ngày. Ngoài ra, máy chủ cung cấp API để gửi dữ liệu cho người dùng thông qua nhiều phương thức khác nhau như một ứng dụng web.

Giao diện người dùng cung cấp dữ liệu môi trường và kết quả dự đoán thông qua ứng dụng web được thiết kế theo tiêu chí đơn giản và thân thiện với người dùng không chuyên.

4.3. Thiết kế chi tiết

Mỗi khối trong 4 khối chính của hệ thống đề được nhóm nghiên cứu thiết kế, xây dựng qua nhiều giai đoạn và nhiều phiên bản khác nhau để phục vụ cho từng trường hợp có đặc điểm riêng. Chi tiết mỗi khối như sau:

4.3.1 Khối các thiết bị quan trắc môi trường

Để có thể thu thập được các thông số đầu vào sử dụng trong mô hình máy học giải quyết bài toán dự đoán mưa, nhóm sử dụng bộ thiết bị SKU:SEN0186 (Hình 5) của DFRobot trên thiết bị quan trắc để thực hiện xây dựng hệ thống cảm biến quan trắc gió và mưa. Các thiết bị bao gồm các thiết bị như hình 6 Các thiết bị trong hình bao gồm:

- Phong kế
- Phong biểu
- Cảm biến mưa
- Board quản lý STC

Được trang bị cảm biến lượng mưa và phong kế, SKU:SEN0186 của DFRobot là một giải pháp phổ biến cho nhiều ứng dụng theo dõi thời tiết với giá cả hợp lý. Thông số cụ thể của SKU:SEN0186 như trong bảng sau:

Thông số kỹ thuật	Điện áp hoạt động: 5V Phạm vi nhiệt độ: -40~80°C Phạm vi độ ẩm: 0~99% Kích thước: 20*18*30 CM Trọng lượng: 4480g
Ứng dụng phổ biến	Trạm khí tượng Quan trắc thời tiết
Giao thức truyền dữ liệu	Serial: 9600bps chu kỳ 1 giây
Định dạng dữ liệu	35 bytes mỗi giây, gồm ký tự kết thúc chuỗi CR/LF Ví dụ: c000s000g000t086r000p000h53b10021\r\n c000 : Hướng gió, 0 độ s000 : Tốc độ gió trung bình (1 phút), 0 dặm/giờ g000 : Tốc độ gió tối đa (5 phút), 0 dặm/giờ t086 : Nhiệt độ, 86 độ Fahrenheit r000 : Lượng mưa trung bình (1 giờ), 0 inches p000 : Lượng mưa tối đa (24 giờ), 0 inches h53 : Độ ẩm, 53% b10021 : Áp suất, 1002.1 hpa

Ngoài ra còn các cảm biến môi trường khác là cảm biến nhiệt độ độ ẩm SHT10 (Hình 7 và 8) của hãng sản xuất SENSIRION và cảm biến áp suất môi trường BMP180 (Hình 9) của hãng sản xuất Bosch.

SHT10 được Sensirion thiết kế với bộ chuyển đổi tương tự sang số 14-bit và mạch giao diện nối tiếp. Nó là một cảm biến loạt chi phí thấp so với các dòng SHT1x khác. Phiên bản được sử dụng trong đề tài này được tinh chỉnh với vỏ bảo vệ cảm biến khỏi các yếu tố môi trường vật lý như nước, bụi, v.v... Một số đặc điểm kỹ thuật của SHT10 như sau:

Đặc điểm kỹ thuật	Dòng sản phẩm: SHT1x
	Độ chính xác % độ ẩm: ± 3
	Độ chính xác $^{\circ}\text{C}$ nhiệt độ: ± 0.4
	Điện áp hoạt động: 2.4V đến 5.5V
	Giao thức truyền dữ liệu: Digital SBus
	Kích thước: 7.5mm x 4.9mm x 2.6mm

BMP180 là một giải pháp cảm biến chi phí thấp để đo áp suất khí quyển và nhiệt độ do Bosch sản xuất. Do thay đổi áp suất theo độ cao, cảm biến này cũng có thể được sử dụng làm máy đo độ cao. Một số đặc điểm kỹ thuật của SHT10 như sau

Đặc điểm kỹ thuật	Điện áp hoạt động: 1.8 ~ 3.6V
	Dòng hoạt động: 0.5uA ở 1Hz
	Giao thức truyền dữ liệu: I2C
	Điện trở kéo lên của I2C được tích hợp
	Tốc độ tối đa trên I2C: 3.5MHz
	Sai số thấp: 0.03hPa (25cm)
	Được tích hợp sẵn mô-đun tinh chỉnh sai số
	Dải áp suất đo được: 300hPa ~ 1100hPa (+9000m đến -500m)
	Trọng lượng: 1.18g
	Kích thước: 21mm x 18mm

Để giao tiếp với Gateway, nhóm trang bị cho các thiết bị quan trắc mô-đun LoRa RFM95W. Mô-đun thu phát LoRa RFM95 do HopeRF sản xuất cung cấp khả năng giao tiếp cực xa với việc giảm thiểu mức tiêu thụ hiện tại. Sử dụng kỹ thuật điều chế Chirp Spread Spectrum, RFM95 có thể giảm độ nhạy xuống $\sim -148\text{dBm}$, gần bằng mức độ nhiễu xung quanh. Kết hợp với Bộ khuếch đại công suất 20dBm, khoảng cách liên lạc có thể đạt tới 15km.

Đặc điểm kỹ thuật	LoRa™ Modem
	168 dB maximum link budget
	+20 dBm - 100 mW constant RF output vs V supply
	+14 dBm high efficiency PA
	Programmable bit rates up to 300 kbps
	High sensitivity: down to -148 dBm
	Bullet-proof front end: IIP3 = -125 dBm
	Excellent blocking immunity
	Low RX current of 103 mA, 200 mA register retention
	Fully integrated synthesizer with a resolution of 61 Hz

	FSK, GFSK, MSK, GMSK, LoRa™ and OOK modulation Built-in bit synchronizer for clock recovery Preamble detection 127 dB Dynamic Range RSSI Automatic RF Sense and CAD with ultra-fast AFC Packet engine up to 256 bytes with CRC Built-in temperature sensor and low battery indicator Package Size : 16*16mm
Ứng dụng	Automated Meter Reading Home and Building Automation Wireless Alarm and Security Systems Industrial Monitoring and Control Long range Irrigation Systems

Ngoài ra, thiết bị sử dụng nguồn nuôi là 2 viên pin li-ion 18650 (Hình 10), dung lượng mỗi viên là 3400mAh. Ngoài ra còn có sử dụng pin mặt trời kích thước 18x18cm công suất 5W của Solarcity (Hình 11) và hình 12) thông qua dụng mạch sạc pin li-ion TP4056 (Hình 13) để duy trì hoạt động của hệ thống.

Các thiết bị quan trắc môi trường được thiết kế dựa trên các tiêu chí sau: Tiết kiệm điện, dễ lắp đặt, bền bỉ cho hoạt động ngoài trời trong thời tiết mưa. Hình 14 thể hiện liên kết giữa các mô-đun trong thiết bị thông qua sơ đồ khối. Có thể thấy được sơ bộ hoạt động của thiết bị thông qua sơ đồ này. Tấm pin năng lượng mặt trời sẽ sạc cho pin vào lúc trời sáng thông qua bộ quản lý nguồn và cung cấp, quản lý năng lượng cho toàn bộ hệ thống. MCU điều khiển các mô-đun khác để thu thập dữ liệu và gửi về máy chủ thông qua Gateway bằng đường truyền LoRa.

Hình 15 cho thấy thiết kế phiên bản đầu tiên sử dụng MCU Atmega328P của hãng Atmel nay thuộc về Microchip Inc. Lý do mà nhóm lựa chọn MCU này là vì được hỗ trợ tốt từ cộng đồng, hệ thống thư viện rộng rãi, giúp nhóm có thể đẩy nhanh tiến độ hoàn thành phần cứng và các thành phần cơ bản để thu thập những dữ liệu mẫu đầu tiên dùng để huấn luyện và xây dựng mô hình máy học.

Sau khi thu thập gần đủ dữ liệu dùng cho việc huấn luyện, xây dựng mô hình máy học, nhóm tiếp hành phát triển một phiên bản khác với nhiều tính năng hơn, tập trung vào tiết kiệm năng lượng và giảm chi phí của thiết bị. Phiên bản mới sử dụng MCU STM8S003K3, đây là một phiên bản trong dòng MCU cơ bản giá của hãng sản xuất STMicroelectronics. Trong thiết kế của phiên bản thứ hai có tích hợp thêm một thành phần mới đó là bộ quản lý sạc pin. Hệ thống chỉ cho phép quá trình sạc pin Li-ion xảy ra khi dung lượng pin còn thấp hơn 35% để giảm tình trạng pin bị chai quá nhanh khi vừa sạc vừa dùng vào ban ngày như ở phiên bản đầu tiên. Hình 16 là hình ảnh thực tế của phiên bản được cải tiến này.

4.3.2 Gateway

Thiết kế gateway bao gồm 2 phiên bản khác nhau: phiên bản đơn giản sử dụng kết nối Wi-Fi và phiên bản sử dụng Raspberry Pi. Mỗi phiên bản có ưu và nhược điểm khác nhau.

Đối với phiên bản đơn giản sử dụng Wi-Fi, tiêu chí thiết kế được đặt lên hàng đầu là chi phí thấp, thiết kế đơn giản và nhỏ gọn. Nó chỉ bao gồm 2 thành phần là mô-đun LoRa RFM95W và mô-đun Wi-Fi ESP8266. Mô-đun ESP8266 vừa có vai trò điều khiển mô-đun LoRa vừa có vai trò quản lý kết nối Internet cũng như Wi-Fi. Tổng giá dưới 10 đô la cho mỗi gateway loại này. Tuy nhiên, nó chỉ hoạt động với kết nối Wi-Fi. Hình 17 và hình 18 cho thấy hình ảnh thực tế của phiên bản gateway này.

Đối với phiên bản Raspberry Pi, nhóm đặt ra mục tiêu thiết kế là khả năng mở rộng của ứng dụng. Trong đó, bán kính vùng vận hành sẽ phụ thuộc vào ăng-ten và thường cao hơn phiên bản ESP8266 sử dụng ăng-ten tích hợp. Ăng-ten được kết nối thông qua tiêu đề IPEX hoặc SMA. Ngoài ra, mỗi gateway phiên bản này sử dụng hai mô-đun LoRa RFM95W, do đó, nó có thể quản lý nhiều thiết bị quan trắc trong một khu vực hơn phiên bản ESP8266 ngay cả trong chế độ multi-channel. Kết nối Internet không giới hạn chỉ có Wi-Fi, nó cũng có thể sử dụng Ethernet. Khả năng mở rộng ứng dụng trong phiên bản này cũng được nâng cao nhờ việc sử dụng hệ điều hành Raspbian, một bản phân phối của nền tảng Linux Debian. Hình 19 cho thấy hình ảnh thực tế của phiên bản gateway này.

Cả hai phiên bản đều hỗ trợ chuẩn LoRaWan và màn hình LCD bên ngoài được cung cấp theo yêu cầu. Màn hình LCD hiện được hỗ trợ là LCD Graphics Nokia5110 với độ phân giải là 84x48 pixel như trong hình 20 và hình 21.

4.3.3 Máy chủ

Hệ thống chủ được thiết kế để xử lý nhiều tác vụ khác nhau như quản lý cơ sở dữ liệu, quản lý kết nối từ các Gateway, cung cấp API lấy dữ liệu hiển thị ra trang web, thực thi dự đoán mưa trên mô hình máy học, thực thi huấn luyện dữ liệu mới cho mô hình máy học. Các tác vụ này được nhóm nghiên cứu chia thành nhiều khối khác nhau, gồm có: khối quản lý cơ sở dữ liệu, khối quản lý kết nối từ các Gateway, khối quản lý mô hình máy học, khối quản lý các trang web. Các khối này liên kết và tương tác với nhau như trong hình 22.

Khối quản lý cơ sở dữ liệu

Thiết kế với mục đích xử lý dữ liệu lớn nên nhóm nghiên cứu đã sử dụng MongoDB làm nền tảng lưu dữ liệu. Cơ sở dữ liệu được cài đặt trên máy chủ, nhóm nghiên cứu thiết kế ứng dụng quản lý truy xuất dữ liệu trên cơ sở dữ liệu này để tránh việc quá tải trên cơ sở dữ liệu khi cho truy cập trực tiếp. Ứng dụng quản lý truy xuất dữ liệu được viết bằng ngôn ngữ Java và sử dụng thư viện tương tác với MongoDB do chính MongoDB Inc cung cấp. Ứng dụng này cho phép các khối khác gián tiếp tương tác với dữ liệu được lưu thông qua kết nối socket nội bộ ở port 5001. Ví dụ: một yêu cầu ghi kết quả dự đoán từ khối quản lý mô hình máy học vào cơ sở dữ liệu có cấu trúc như

sau: {"code":2, "nodeId":"Weather_Node_1", "date":"2018-11-10 13:15", "rain":1, "AmountOfRain":20}.

Cấu trúc của khối này cũng được thiết kế đặc biệt giúp giải thuật xử lý các yêu cầu không làm hệ thống bị lỗi khi có quá nhiều yêu cầu được gửi đến cùng lúc. Cấu trúc này có thể được thể hiện như trong hình 23. Trong cấu trúc này, Workload chính là phần công việc đã được giải mã từ chuỗi yêu cầu được gửi đến thông qua socket và được xếp vào hàng đợi chờ được xử lý theo nguyên tắc First in First out, RequestHandler là đối tượng xử lý các Workload được lưu trong hệ thống. Hiện tại, trong toàn bộ hệ thống chỉ được tạo 4 RequestHandler chạy trên 4 tiến trình độc lập và song song nhau. Việc tạo ra nhiều RequestHandler hơn sẽ tiêu tốn nhiều tài nguyên hệ thống hơn như bù lại sẽ giảm thời gian chờ của các yêu cầu hơn khi có nhiều yêu cầu đến cùng lúc. Nhóm nghiên cứu lựa chọn số lượng RequestHandler là 4 cũng đã qua nhiều bước thử nghiệm độ trễ của yêu cầu bằng phương pháp "Stress Test".

Khối quản lý kết nối từ các Gateway

Giống như tên của nó, khối này có chức năng chính là quản lý các kết nối và gói tin được chuyển từ các gateway. Khối này được thiết kế trên tiêu chí là chia sẻ thông tin và có thể mở rộng. Tiêu chí mở rộng ở đây mà nhóm muốn hướng tới chính là việc mở rộng mô hình máy học, cụ thể là xử dụng nhiều dữ liệu đầu vào được thu thập từ nhiều thiết bị quan trắc quanh các khu vực thay vì chỉ sử dụng dữ liệu từ một thiết bị quan trắc như hiện tại.

Khối quản lý kết nối từ các Gateway được thiết kế sử dụng ngôn ngữ NodeJS là ngôn ngữ phát triển chính và kết nối đến khối quản lý cơ sở dữ liệu thông qua kết nối socket để ghi gói tin nhận được nếu như gói tin đó hợp lệ. Các gateway kết nối đến khối này thông qua giao thức MQTT, một giao thức được tạo ra nhằm chia sẻ thông tin giữa các thiết bị như một mạng xã hội. Các gói tin gateway nhận được từ các thiết bị quan trắc sẽ được chuyển tiếp đến khối quản lý kết nối từ các Gateway này, ở đây các gói tin sẽ được xử lý thêm một lần nữa để chắc chắn rằng không có lỗi trong gói tin trong quá trình truyền nhận. Cuối cùng, khối quản lý kết nối từ các Gateway sẽ tạo ra một yêu cầu ghi dữ liệu nhận được vào cơ sở dữ liệu, sử dụng cho việc dự đoán mưa cũng như các phát triển sau này.

Khối quản lý mô hình máy học

Chức năng chính của khối này là quản lý mô hình máy học, lập lịch và kiểm soát quá trình học của mô hình, kích hoạt mô hình để đưa ra dự đoán mưa. Khi thiết kế mô hình này nhóm nghiên cứu đã khảo sát và đưa ra bảng so sánh ưu nhược điểm của các mô hình máy học như sau:

Thuật toán	Ưu điểm	Nhược điểm
Naïve Bayes	+ Dễ dàng cài đặt. + Thời gian thi hành nhanh + Đạt kết quả tốt trong phần lớn các trường hợp	+ Giả thiết về tính độc lập điều kiện của các thuộc tính làm giảm độ chính xác

Binary Classifiers	<ul style="list-style-type: none"> + Không cần có giả thiết dữ liệu hai class là linearly separable. + Dễ áp dụng vào bài toán. + Thời gian thi hành nhanh. + Thích hợp dùng cho bài toán phân lớp 	<ul style="list-style-type: none"> + Chỉ phù hợp với loại dữ liệu mà hai class là gần với linearly separable. + Không làm việc được với kiểu dữ liệu phi tuyến. + Yêu cầu các điểm dữ liệu được tạo ra một cách độc lập với nhau.
K-nearest neighbors	<ul style="list-style-type: none"> + Độ phức tạp tính toán của quá trình training là bằng 0. + Việc dự đoán kết quả của dữ liệu mới rất đơn giản. + Không cần giả sử gì về phân phối của các class 	<ul style="list-style-type: none"> + KNN rất nhạy cảm với nhiễu khi K nhỏ. + Việc tính khoảng cách tới từng điểm dữ liệu trong training set sẽ tốn rất nhiều thời gian, đặc biệt là với các cơ sở dữ liệu có số chiều lớn và có nhiều điểm dữ liệu. + Việc lưu toàn bộ dữ liệu trong bộ nhớ cũng ảnh hưởng tới hiệu năng của KNN.
Softmax Regresion	<ul style="list-style-type: none"> + Có thể giải quyết Các bài toán classification có nhiều class. + Sử dụng tốt trong các mạng Neural có nhiều lớp. Thường được ứng dụng ở lớp cuối cùng của mô hình. + Không cần có giả thiết dữ liệu các class là linearly separable. 	<ul style="list-style-type: none"> + Chỉ phù hợp với loại dữ liệu mà các class là gần với linearly separable.
Multi-layer Perceptron and Back-propagation	<ul style="list-style-type: none"> + Không giới hạn dữ liệu thuộc linear hay không. + Boundary phụ thuộc vào dữ liệu train. + Đạt kết quả tốt trong phần lớn các trường hợp. + Có thể mở rộng, kết hợp thêm các thuật toán khác để đạt được kết quả mong muốn. 	<ul style="list-style-type: none"> + Gặp khó khăn khi dữ liệu rời rạc. + Tốn nhiều thời gian để tìm ra được số lượng units, hidden layers thích hợp. + Khi số lượng hidden layers lớn lên, số lượng hệ số cần tối ưu cũng lớn lên và mô hình sẽ trở nên phức tạp. Dẫn đến tốc độ chậm đi. Dẫn đến hiện tượng Overfitting.

Support Vector Machine	<ul style="list-style-type: none"> + Không cần xác định mô hình của đối tượng như neuron, fuzzy logicm, ... + Giải quyết tốt bài toán có dữ liệu có số chiều lớn. + Giải quyết vấn đề overfitting rất tốt. + Phân lớp nhanh, hiệu suất tổng hợp tốt, hiệu suất tính toán cao, tiết kiệm bộ nhớ, tính linh hoạt 	<ul style="list-style-type: none"> + Độ phức tạp cao. + Hạn chế khi xử lý dữ liệu kiểu số. + Dễ SVM hiệu quả thì dữ liệu cần phải được trích-chọn các thuộc tính phù hợp, SVM không thể lựa chọn thuộc tính được nên phải tự lựa chọn thuộc tính hoặc dùng các thuật toán khác. + Trong trường hợp số lượng thuộc tính (p) của tập dữ liệu lớn hơn rất nhiều so với số lượng dữ liệu (n) thì SVM cho kết quả khá tồi. + Chưa thể hiện rõ tính xác suất
------------------------	--	---

Dựa vào bảng so sánh trên, nhóm đã quyết định xây dựng khối này bằng giải thuật Multi-layer Perceptron and Back-propagation với ngôn ngữ C/C++ và trải qua nhiều giai đoạn thiết kế như trong bảng sau:

Bước 1	Quá trình lựa chọn các biến
Bước 2	Quá trình thu thập dữ liệu
Bước 3	Xử lý và phân tích dữ liệu
Bước 4	Phân tích dữ liệu cho từng giai đoạn: Huân luyện, kiểm tra, công nhận
Bước 5	Lựa chọn mô hình phù hợp Tính toán số lượng lớp ẩn Tính toán số lượng neuron ẩn Tính toán số lượng đầu ra Xác định loại hàm truyền
Bước 6	Xác định tiêu chuẩn đánh giá kết quả

Bước 7	Lựa chọn kiểu huấn luyện
Bước 8	Tiến hành thực hiện mô hình

Hoạt động của khối này được thể hiện một cách tổng quan chi tiết trong hình 24, trong đó mô hình máy học được thể hiện chi tiết trong hình 25. Chi tiết hoạt động của toàn bộ mô hình có thể được tóm tắt như sau:

Bước 1	Nhận inputs đầu vào
Bước 2	Tính trọng số ở lớp ẩn theo hàm truyền sigmoid
Bước 3	Tính trọng số ở lớp output theo hàm truyền softmax
Bước 4	Tính toán sai số và hiệu chỉnh lại trọng số dựa trên thuật toán back-propagation
Bước 5	Nêu số lượng inputs để train chưa hết thì quay lại Bước 1
Bước 6	Nếu chưa thực hiện đủ số bước lặp thì quay lại Bước 1 với inputs ban đầu

Khối quản lý các trang web

Khối quản lý các trang web chịu trách nhiệm xử lý các tác vụ liên quan đến web như một web server, cụ thể lấy dữ liệu môi trường, kết quả dự đoán, thông tin về thiết bị quan trắc sau đó hiển thị lên web. Được phát triển bằng ngôn ngữ NodeJS và cũng như các khối khác, khối này liên kết với khối quản lý cơ sở dữ liệu thông qua giao thức socket để lấy các dữ liệu cần thiết.

4.3.4 Giao diện người dùng

Back-end được phát triển bằng ngôn ngữ NodeJS và được trình bày ở mục 4.3.4 phần Khối quản lý các trang web, và front-end thiết kế bằng HTML5/CSS với tiêu chí đơn giản và thân thiện với người dùng. Một số hình ảnh thực tế của hệ thống các trang web ở các hình 26, 27, 28.

4.4. Tích hợp hệ thống và kiểm thử

Hoàn thành được hệ thống

Hình 29 là hình chụp của trạm thu dữ liệu. Thiết bị quan trắc gồm các cảm biến và mạch điều khiển được đặt trong hộp để tránh mưa ướn. Hoạt động của thiết bị quan trắc ổn định trong thời gian dài, nhóm đã lắp đặt thiết bị từ tháng 7/2018 đến tháng 11/2018

ở ngoài trời nhưng vẫn hoạt động ổn định. Dữ liệu thu thập được gửi tới máy chủ mỗi 5 phút một lần.

Pin năng lượng mặt trời kết hợp với bộ sạc với mức năng lượng mặt trời cung cấp đủ nguồn cho tất cả các thiết bị của hệ thống và đồng thời sạc đầy 2 viên pin li-ion, có thể cung nguồn cho các thiết bị khi thời gian không có ánh sáng mặt trời.

Hệ thống máy chủ cũng được cài đặt tại máy chủ VPS ở trường Đại học Công nghệ Thông tin hoạt động và gateway được lắp đặt tại đây kết nối vào máy chủ qua đường truyền Wi-Fi cũng rất ổn định.

Độ chính xác dự đoán

Để xác định được tỉ lệ của việc dự đoán, nhóm tiến hành thu thập giá trị và dự đoán kết quả để xác định. Nhóm xác định tỉ lệ dự đoán được mưa dự đoán được thời lượng xảy ra mưa.

Để có thể có được dữ liệu của thời tiết, nhóm thực hiện thu thập dữ liệu từng ngày, ghi chép thời lượng và thời gian mưa thực tế, sau đó nhập thời lượng và thời gian mưa vào hệ thống. Thông qua các quá trình thống kê và tính toán, nhóm xác định được tỉ lệ dự đoán của hệ thống.

Trải qua 3 giai đoạn phát triển của hệ thống, nhóm đã ghi lại kết quả dự đoán. Sau đây là kết quả một số ngày được dự đoán ở từng giai đoạn phát triển, kết quả được lựa chọn hoàn toàn ngẫu nhiên:

Giai đoạn 1: từ 25/08/2018 đến 15/09/2018

Kết quả: kết quả dự đoán được sau khi train trên mô hình neural không cho kết quả khả thi, những dự đoán cho kết quả rất không chính xác.

Dữ liệu vào mô hình máy học:

- Tốc độ gió, hướng gió: Căn cứ vào kiểm thức địa lý sách giáo khoa, hương gió thổi từ biển đông mang theo hơi nước, mây mưa, ...
- Giờ trong ngày: môi trường ở thời điểm khác nhau là khác nhau. Tùy thuộc mỗi mùa, tháng thời điểm mưa trong ngày là khác nhau. VD: tháng 10 mưa thường tập trung vào khoảng 6 giờ đến 10 giờ sáng
- Nhiệt độ thấp, cao nhất trong 1 giờ: Khoảng nhiệt độ (thấp nhất, cao nhất), mà tại đây ta có thể xác định mưa hay không có mưa. Do không thể nào xác định chính xác được là tại 1 điểm nhiệt độ chính xác mà có thể dự đoán được nên ta lấy khoảng.
- Độ ẩm thấp, cao nhất trong 1 giờ: Khoảng độ ẩm (thấp nhất, cao nhất), mà tại đây ta có thể xác định mưa hay không có mưa. Do không thể nào xác định chính xác được là tại 1 điểm độ ẩm chính xác mà có thể dự đoán được nên ta lấy khoảng.
- Tổng độ giảm áp suất trong 2 giờ

Nguyên nhân và cải tiến:

- Đầu ra mong muốn không thích hợp với mô hình, phương hướng cải tiến là đầu ra: nhận kết quả là 0 và 1, chia bộ dữ liệu thành 2 lớp là có mưa, không có mưa.
- Đầu vào nhiệt độ thấp nhất, nhiệt độ cao nhất, độ ẩm thấp nhất, độ ẩm cao nhất. Không dùng được, do không giúp bộ dữ liệu phân chia thành 2 lớp rõ ràng, không thích hợp với phương án mới. Cải tiến: nhiệt độ trung bình trong 1 giờ, độ ẩm trung bình trong 1 giờ.

Giai đoạn 2: từ 25/09/2018 đến 15/10/2018

Ngày	Giờ trong ngày	Kết quả thực tế đo được	Kết quả dự đoán	Chính xác	Độ chính xác
28/09/2018	2	Không mưa	Không mưa	Ok	42%
	3	Không mưa	Không mưa	Ok	
	4	Không mưa	Không mưa	Ok	
	5	Có mưa	Không mưa	No	
	6	Có mưa	Không mưa	No	
	7	Có mưa	Không mưa	No	
	8	Có mưa	Không mưa	No	
01/10/2018	12	Không mưa	Có mưa	No	50%
	13	Không mưa	Có mưa	No	
	14	Không mưa	Có mưa	No	
	15	Không mưa	Không mưa	Ok	
	16	Không mưa	Không mưa	Ok	
	17	Không mưa	Không mưa	Ok	
	18	Không mưa	Không mưa	OK	
	19	Không mưa	Không mưa	Ok	
	20	Có mưa	Không mưa	No	
	21	Có mưa	Không mưa	No	
	22	Có mưa	Không mưa	No	
	23	Có mưa	Có mưa	Ok	
03/10/2018	10	Có mưa	Có mưa	Ok	50%
	11	Có mưa	Có mưa	Ok	
	12	Có mưa	Có mưa	Ok	
	13	Có mưa	Không mưa	No	
	14	Có mưa	Không mưa	No	
	15	Có mưa	Không mưa	No	
	16	Không mưa	Không mưa	Ok	
	17	Không mưa	Không mưa	Ok	
	18	Không mưa	Không mưa	Ok	
	19	Không mưa	Có mưa	No	

	20	Không mưa	Có mưa	No	
	21	Không mưa	Có mưa	No	

Đầu vào:

- Tốc độ gió: không thay đổi
- Hướng gió: chia thành 12 hướng tương ứng (mỗi hướng 30 độ).
- Giờ trong ngày: không thay đổi.
- Nhiệt độ trung bình trong 1 giờ: đặc trưng nhiệt độ
- Độ ẩm trung bình trong 1 giờ: đặc trưng độ ẩm
- Tổng độ giảm áp suất trong 2 giờ: không thay đổi

Đầu ra: nhận giá trị 0 và 1

Mô hình sử dụng ở giai đoạn này cũng đã được chỉnh sửa, thêm giải thuật Backpropagation và thay đổi nội dung hàm truyền ở các lớp ẩn thành Sigmoid và ở lớp đầu ra thành Softmax Regression.

Kết quả thực tế đo được có độ sai lệch so với kết quả ngoài thực tế, cách khắc phục là tăng giá trị khi thu thập được từ cảm biến mưa xác định mưa lên để loại bỏ trường hợp chỉ mưa nhẹ vài hạt nhưng vẫn nhận đó là đã mưa.

Nhận xét: Kết quả dự đoán cho kết quả không cao, chưa thỏa yêu cầu, cần cải thiện.

Cải thiện:

- Nhiệt độ trung bình trong 1 giờ -> Độ biến thiên nhiệt độ trong 1 giờ.
- Độ ẩm trung bình trong 1 giờ -> Độ biến thiên độ ẩm trong 1 giờ.
- Thể hiện rõ hơn đặc trưng, sự thay đổi của nhiệt độ, độ ẩm. Giúp phân rõ hơn 2 lớp trong dữ liệu.

Giai đoạn 3: từ 20/10/2018 đến 31/11/2018

Ngày	Giờ trong ngày	Kết quả thực tế đo được	Kết quả dự đoán	Chính xác	Độ chính xác
03/10/2018	2	Không mưa	Không mưa	Ok	87%
	5	Không mưa	Không mưa	Ok	
	10	Không mưa	Không mưa	Ok	
	12	Có mưa	Không mưa	No	
	15	Có mưa	Có mưa	Ok	
	18	Không mưa	Không mưa	Ok	
	23	Không mưa	Không mưa	Ok	

07/10/2018	2	Không mưa	Không mưa	Ok	90%
	5	Không mưa	Không mưa	Ok	
	6	Có mưa	Không mưa	No	
	8	Không mưa	Không mưa	Ok	
	13	Không mưa	Không mưa	Ok	
	17	Không mưa	Không mưa	Ok	
	19	Không mưa	Không mưa	Ok	
	22	Không mưa	Không mưa	Ok	
18/10/2018	2	Không mưa	Không mưa	Ok	78%
	7	Không mưa	Không mưa	Ok	
	8	Có mưa	Có mưa	Ok	
	9	Có mưa	Không mưa	No	
	10	Có mưa	Không mưa	No	
	12	Có mưa	Có mưa	Ok	
	15	Không mưa	Không mưa	Ok	
	19	Không mưa	Không mưa	Ok	
	20	Không mưa	Không mưa	Ok	

Ở giải đoạn này, nhóm đã chỉnh sửa thêm giải thuật Dropout vào mô hình máy học để kiểm soát quá trình huấn luyện của mô hình máy học. Dropout là cách thức mà chúng ta giả định một phần các đơn vị bị ẩn đi trong quá trình huấn luyện, qua đó làm giảm tích hòa trộn (hay nói cách khác là 1 đơn vị ẩn không thể dựa vào 1 đơn vị khác để sửa lỗi của nó, để cho chúng ta thấy các đơn vị ẩn không đáng tin cậy). Tại mỗi bước trong quá trình huấn luyện, khi thực hiện Forward Propagation (Lan truyền xuôi) đến lớp sử dụng Dropout, thay vì tính toán tất cả đơn vị có trên lớp, tại mỗi đơn vị ta "giao xúc xắc" xem đơn vị đó có được tính hay không dựa trên xác suất.

4.5. Kết quả nghiên cứu

Nhìn chung nghiên cứu đã đạt được các tiêu chí đề ra ban đầu như xây dựng thành công hệ thống dự đoán mưa trong khu vực nhỏ, hoạt động độc lập và ổn định trong thời tiết mưa bão; ứng dụng công nghệ máy học để cải thiện độ chính xác của dự đoán theo thời gian sử dụng hệ thống; sử dụng công nghệ truyền dữ liệu LoRa để áp dụng vào việc truyền nhận tín hiệu, giúp gia tăng khoảng cách lắp đặt hệ thống. Kết quả về độ chính xác của dự đoán cũng rất khả quan khi đạt được đến gần 90% tỉ lệ dự đoán mưa chính xác.

Nhưng xen vào đó vẫn còn tồn đọng một số điểm chưa đạt được như việc số lượng thiết bị quan trắc tăng lên kéo theo các hoạt động trên băng tần của LoRa cũng dày đặc hơn, việc này dẫn đến tỉ lệ gói tin bị lỗi hay mất gói tin cũng tăng lên do đụng độ. Để giải quyết vấn đề này cần có một giải thuật giải quyết đụng độ. Qua khảo sát sơ bộ, nhóm nghiên cứu thấy được các giải thuật giải quyết đụng độ cơ bản và phổ biến hiện nay chỉ có Pure ALOHA và Slotted ALOHA, cả hai tuy giải quyết được phần nào tình trạng đụng độ gói tin nhưng tỉ lệ vẫn còn cao. Thêm vào đó, hệ thống các thiết bị quan trắc đều gửi gói tin theo một chu kỳ nhất định nên nhóm đề xuất hướng giải quyết phát triển một giải thuật đơn giản đồng bộ và lập lịch việc gửi gói tin của mỗi thiết bị quan trắc. Tuy nhiên, do giới hạn về thời gian của đề tài nên nhóm đã đưa việc này vào hướng phát triển.

Hướng phát triển:

- Thiết kế thiết bị quan trắc nhỏ gọn hơn, thẩm mỹ và tiết kiệm năng lượng hơn
- Dựa trên lý thuyết dự báo thời tiết và nghiên cứu thực nghiệm, áp suất giảm mạnh dẫn đến mưa trước 4 giờ. Tuy nhiên, để dự báo lượng mưa trong hơn 4 giờ ở giai đoạn nâng cao, cần có hình ảnh của đám mây.
- Tăng số lượng nút giám sát môi trường để đảm bảo rằng dữ liệu được thu thập là chính xác và đầy đủ
- Phát triển hệ thống thành mạng để có thể theo dõi thời tiết của một khu vực lớn hơn
- Mở rộng mô hình máy học, thêm dữ liệu đầu vào gồm hình ảnh mây và dữ liệu của các thiết bị quan trắc ở các khu vực lân cận
- Thêm cập nhật chương trình cơ sở OTA cho Node và Gateway để cải thiện khả năng sử dụng và giảm chi phí nâng cấp chương trình cơ sở thủ công
- Thêm kết nối Ethernet vào Cổng, đảm bảo khả năng sử dụng cho người dùng
- Phát triển hệ thống Chẩn đoán lỗi phần cứng, cải thiện độ ổn định của phần cứng, đặc biệt là các nút giám sát môi trường khi hoạt động liên tục ngoài trời và hệ thống sẽ thông báo khi phần cứng gặp sự cố

5. Tên sản phẩm

Thiết bị quan trắc, dự báo thời tiết phạm vi nhỏ

6. Hiệu quả, phương thức chuyển giao kết quả nghiên cứu và khả năng áp dụng

Đề tài mang tính mới ở thị trường trong nước, giải quyết được nhu cầu cụ thể là dự đoán thời tiết hay chi tiết hơn là mưa trong phạm vi nhỏ. Ứng dụng được công nghệ máy học giúp hệ thống dễ dàng tương thích được với môi trường đa dạng hơn mà không cần thay đổi thuật toán hay phần cứng.

Thiết bị được đóng gói tương đối tốt bằng hộp nhựa và gia cố thêm bằng khung sắt chắc chắn, giúp thiết bị hoạt động ổn định ngoài trời, dưới thời tiết mưa bão. Tuy nhiên, để có thể hướng đến mục đích thương mại sản phẩm, thiết bị nên có giải pháp đóng gói tốt hơn, tối thiểu hóa khả năng mạch điện bên trong bị ẩm hoặc ướt do thời tiết.

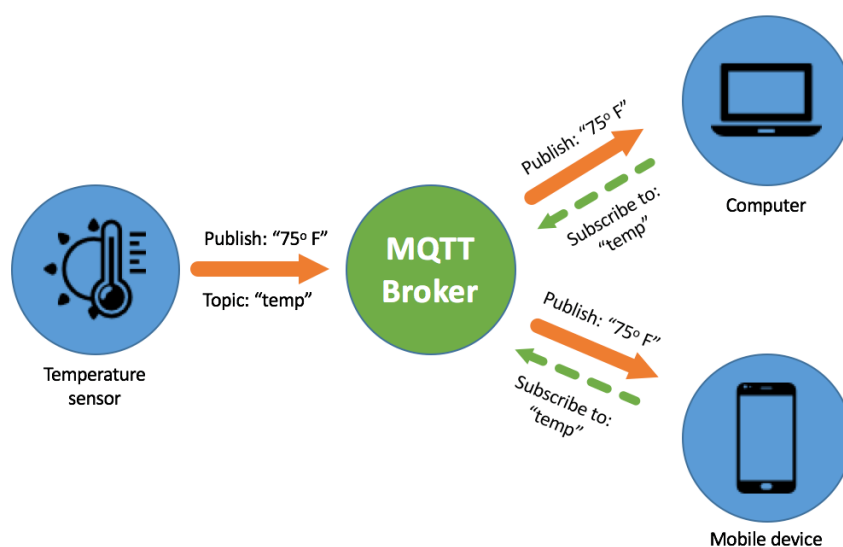
Đề tài mang tính ứng dụng cao, nghiên cứu chế tạo sản phẩm dự đoán lượng mưa trong khu vực nhỏ, giải quyết được các nhu cầu dự đoán thời tiết cụ thể. Thêm vào đó, việc ứng dụng công nghệ máy học vào lĩnh vực dự đoán thời tiết cũng là bước đệm đưa công nghệ mới này vào nhiều lĩnh vực hơn.

Khả năng thương mại hóa sản phẩm ở mức tương đối, do mục đích sử dụng của thiết bị nhắm đến nhóm đối tượng cụ thể. Sản phẩm còn giải quyết được vấn đề dự đoán thời tiết trong phạm vi nhỏ khi trên thị trường còn ít sản phẩm đối thủ cạnh tranh. Ngoài ra, việc thích nghi với môi trường và cải thiện độ chính xác theo thời gian hoạt động cũng là một ưu thế của sản phẩm. Tuy nhiên, cần phải nâng cao tính hoàn thiện của sản phẩm để có thể tăng độ hấp dẫn cũng như giảm khả năng hỏng và chi phí vận hành.

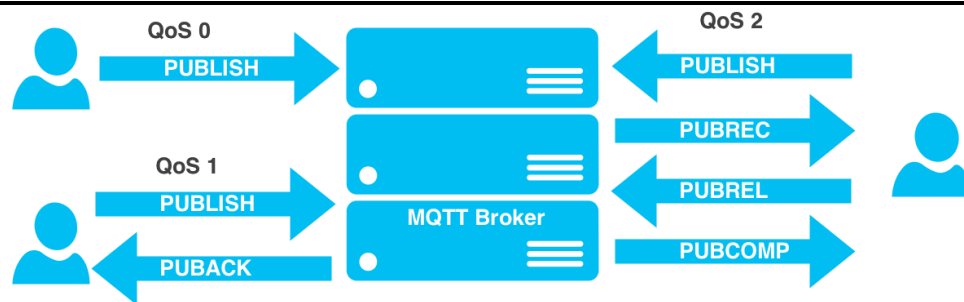
7. Hình ảnh, sơ đồ minh họa chính

	Local Area Network Short Range Communication	Low Power Wide Area (LPWAN) Internet of Things	Cellular Network Traditional M2M
	40%	45%	15%
😊	Well established standards In building	Low power consumption Low cost Positioning	Existing coverage High data rate
😞	Battery Live Provisioning Network cost & dependencies	High data rate Emerging standards	Autonomy Total cost of ownership
	Bluetooth 4.2	LoRa	GSMA 3G+ / H+ 4G

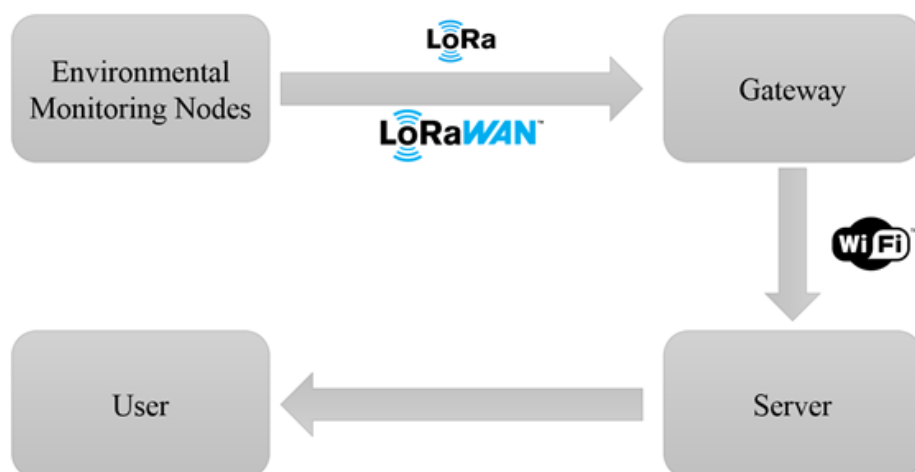
Hình 1: Bảng so sánh ưu/nhược điểm của các công nghệ truyền dữ liệu không dây



Hình 2: Cấu trúc của một hệ thống MQTT cơ bản



Hình 3: Các khái niệm QoS trong MQTT



Hình 4: Sơ đồ khối hệ thống



Hình 5: Bộ thiết bị quan trắc thời tiết DFRobot SKU:SEN0186



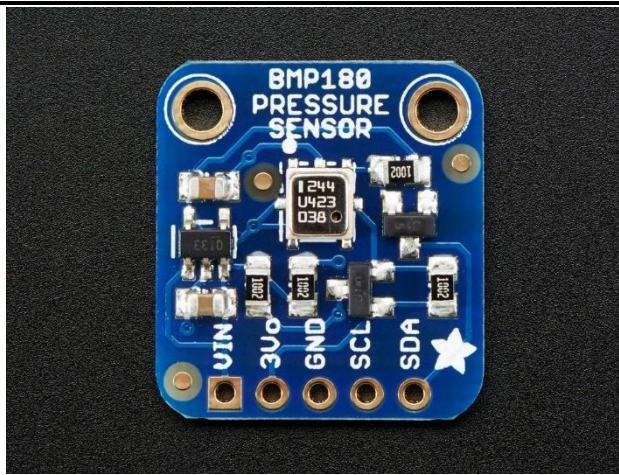
Hình 6: Các bộ phận của bộ thiết bị quan trắc thời tiết DFRobot SKU:SEN0186



Hình 7: Cảm biến nhiệt độ/độ ẩm SHT10



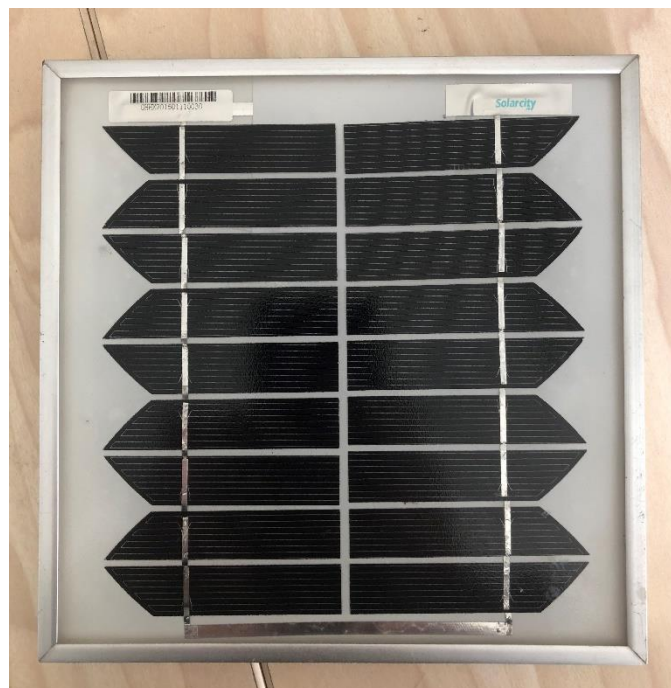
Hình 8: Cảm biến nhiệt độ/độ ẩm SHT10 tháo đầu bảo vệ



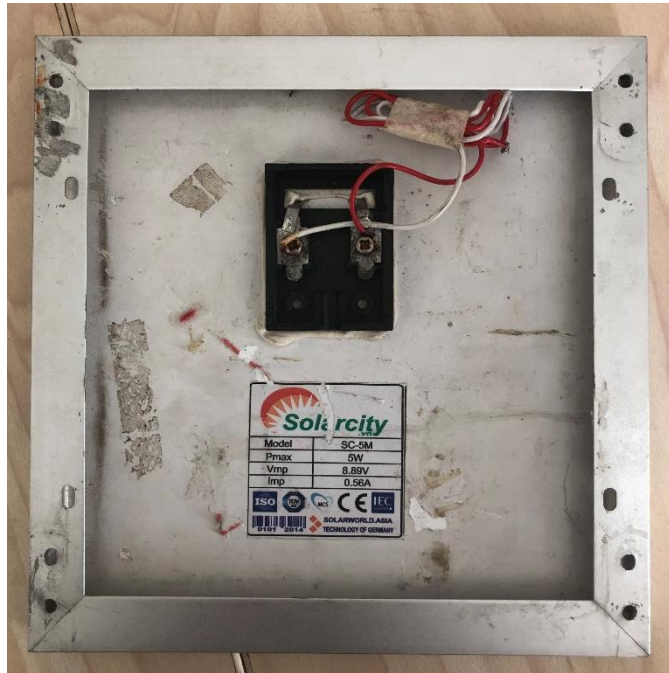
Hình 9: Cảm biến áp suất môi trường BMP180



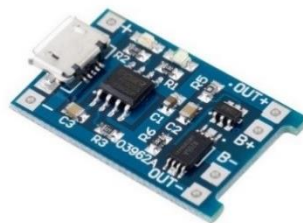
Hình 10: Hai viên pin Panasonic NCR18650



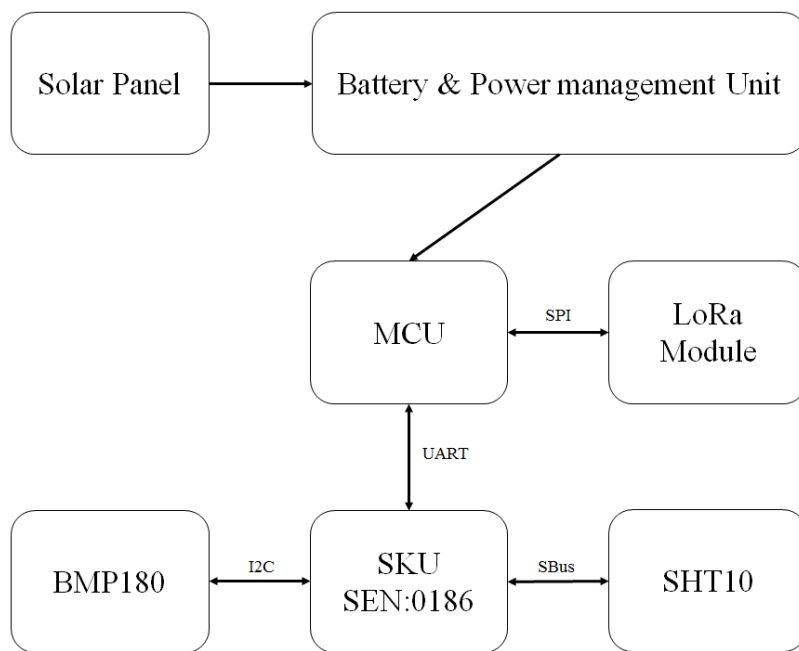
Hình 11: Mặt trước pin mặt trời kích thước 18x18cm công suất 5W của Solarcity



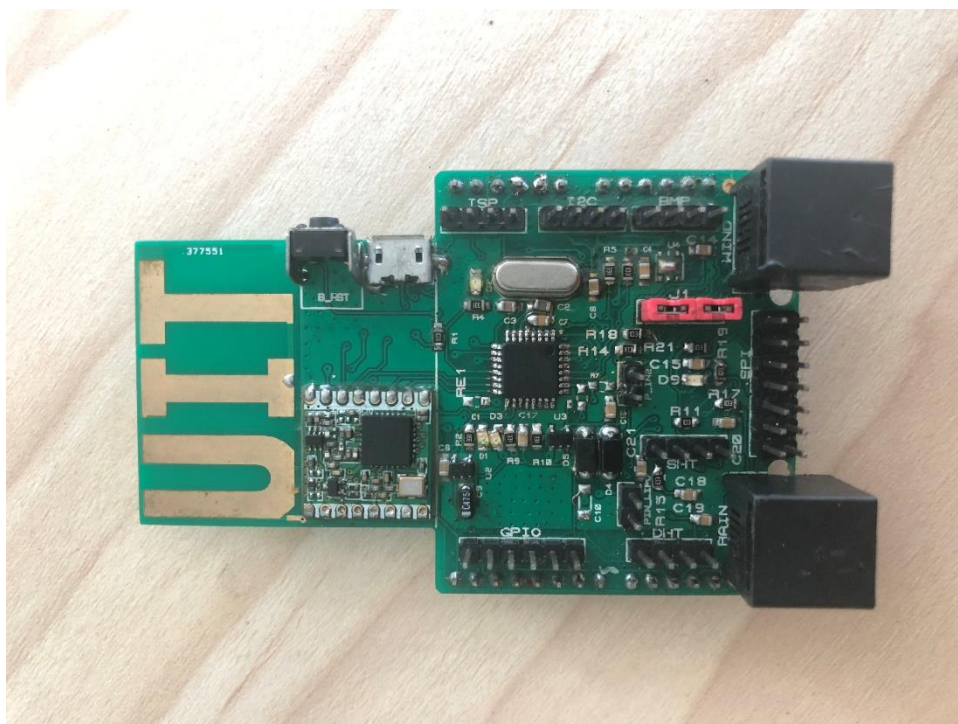
Hình 12: Mặt sau pin mặt trời kích thước 18x18cm công suất 5W của Solarcity



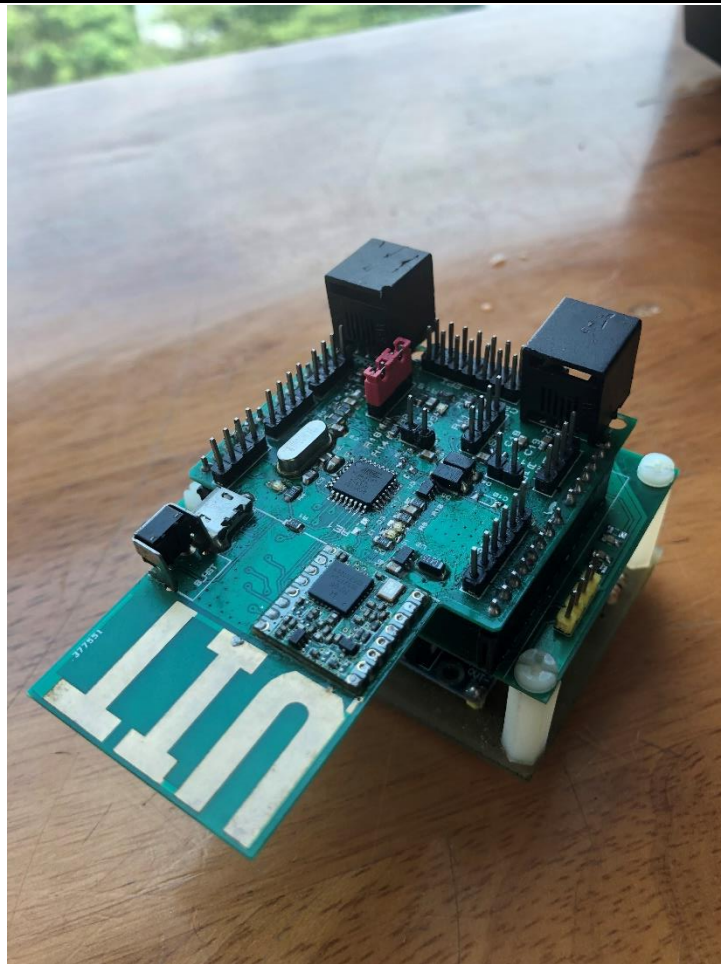
Hình 13: Mạch sạc pin Li-ion TP4056



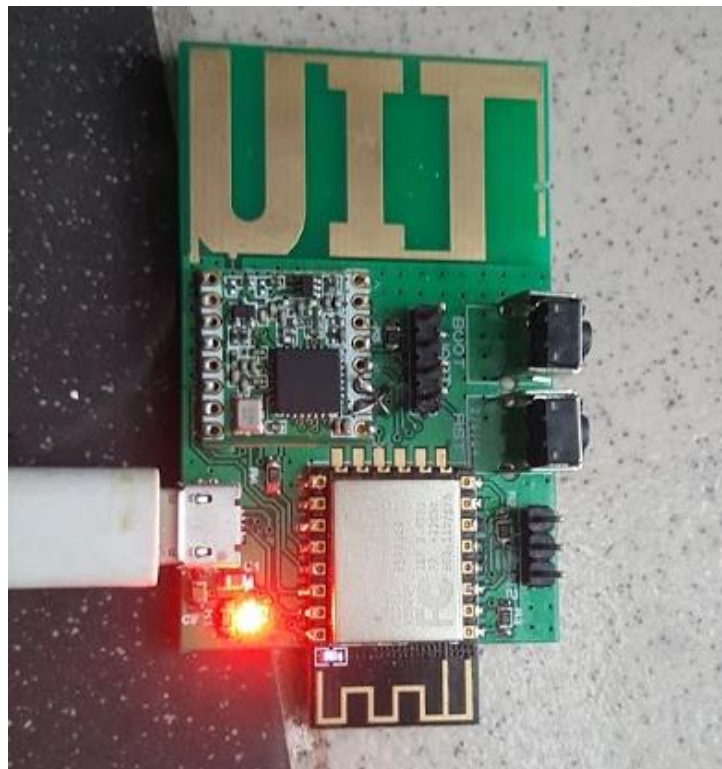
Hình 14: Sơ đồ khối liên kết giữa các mô-đun trong thiết bị quan trắc môi trường



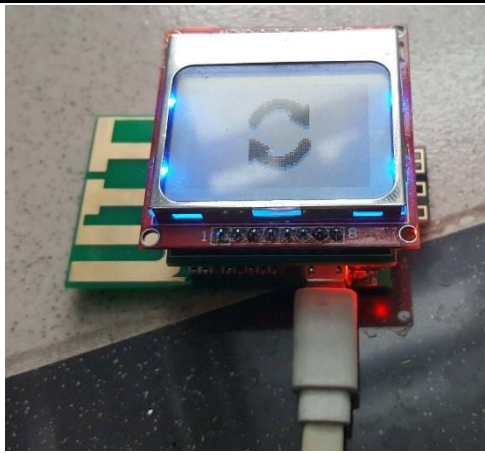
Hình 15: Hình ảnh thực tế của phiên bản thiết bị quan trắc đầu tiên sử dụng MCU Atmega328P



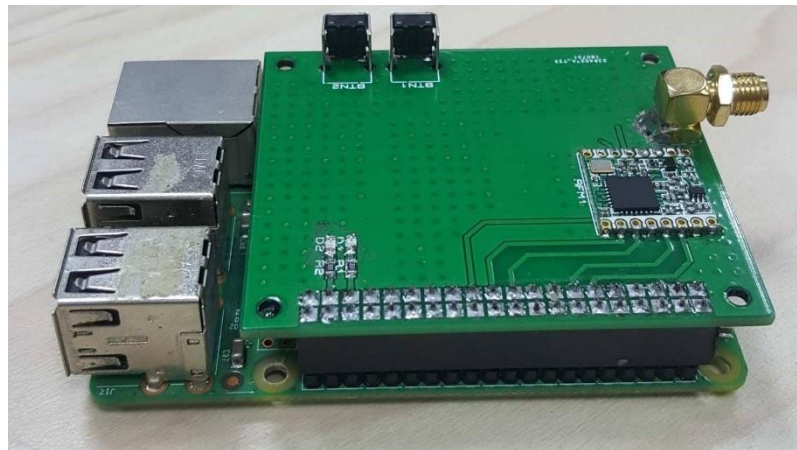
Hình 16: Hình ảnh thực tế của phiên bản thiết bị quan trắc cải tiến sử dụng MCU STM8



Hình 17: Gateway phiên bản 1, sử dụng Wi-Fi



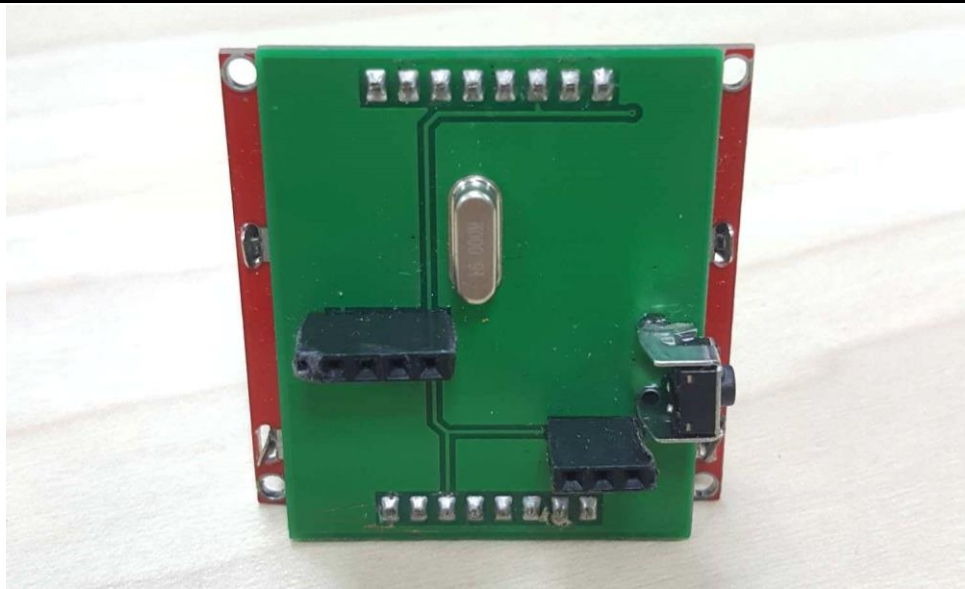
Hình 18: Gateway phiên bản 1 với LCD



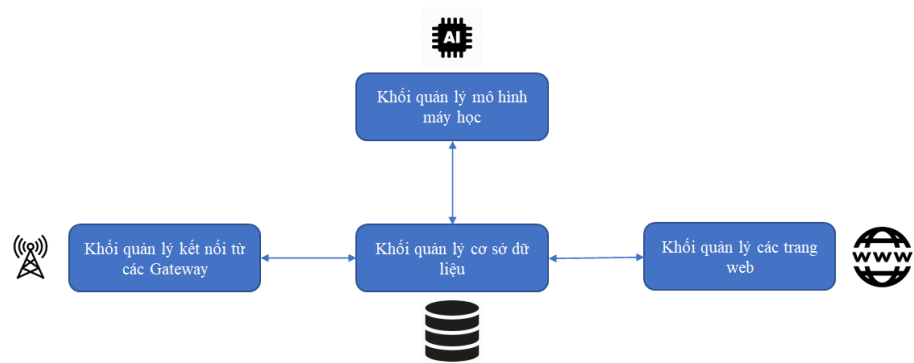
Hình 19: Gateway phiên bản 2 với Raspberry Pi 3



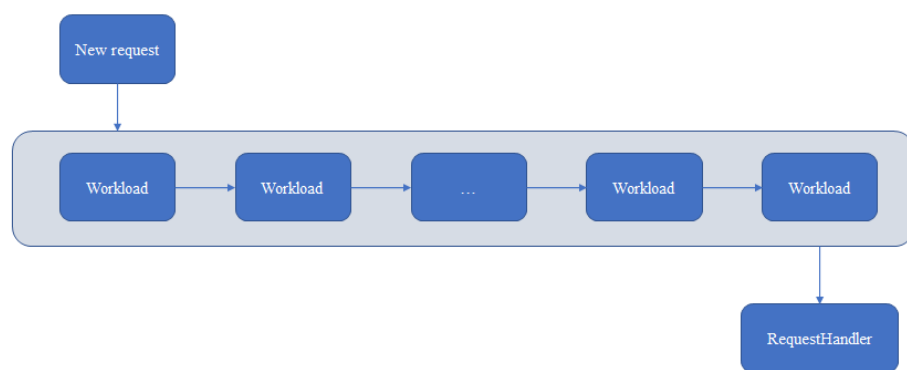
Hình 20: Mặt trước Nokia LCD5110



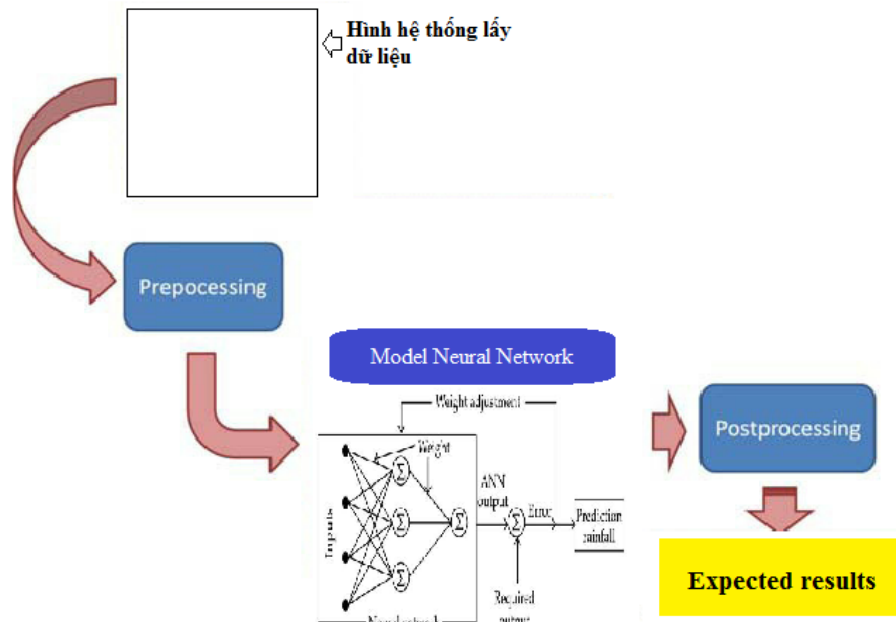
Hình 21: Mặt sau Nokia LCD5110



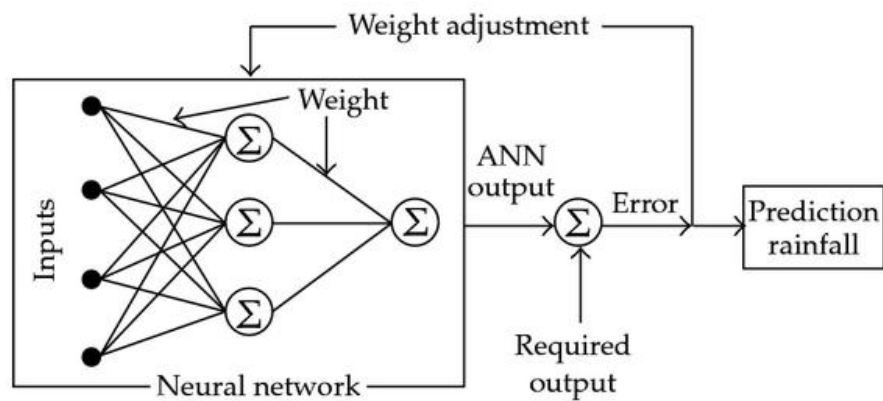
Hình 22: Sơ đồ khối các mô-đun trong máy chủ



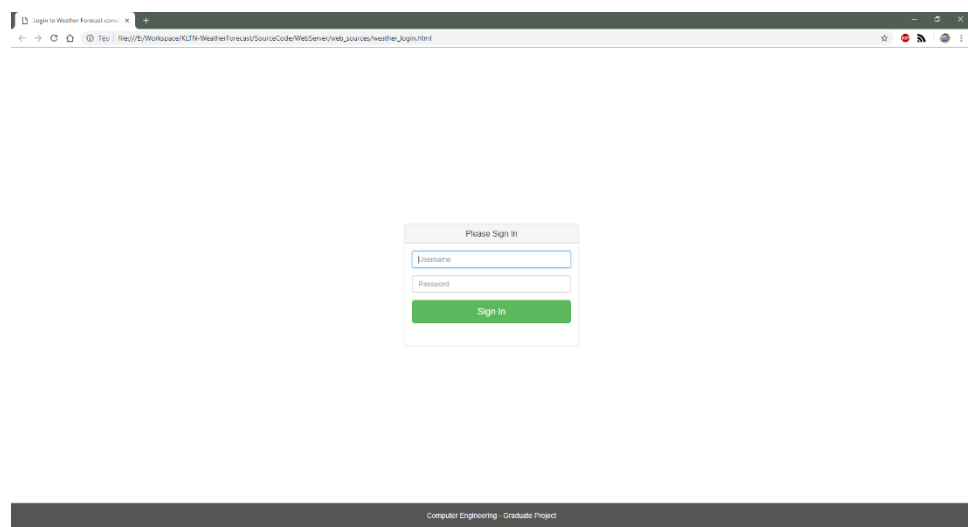
Hình 23: Sơ đồ đường đi dữ liệu trong khối quản lý cơ sở dữ liệu



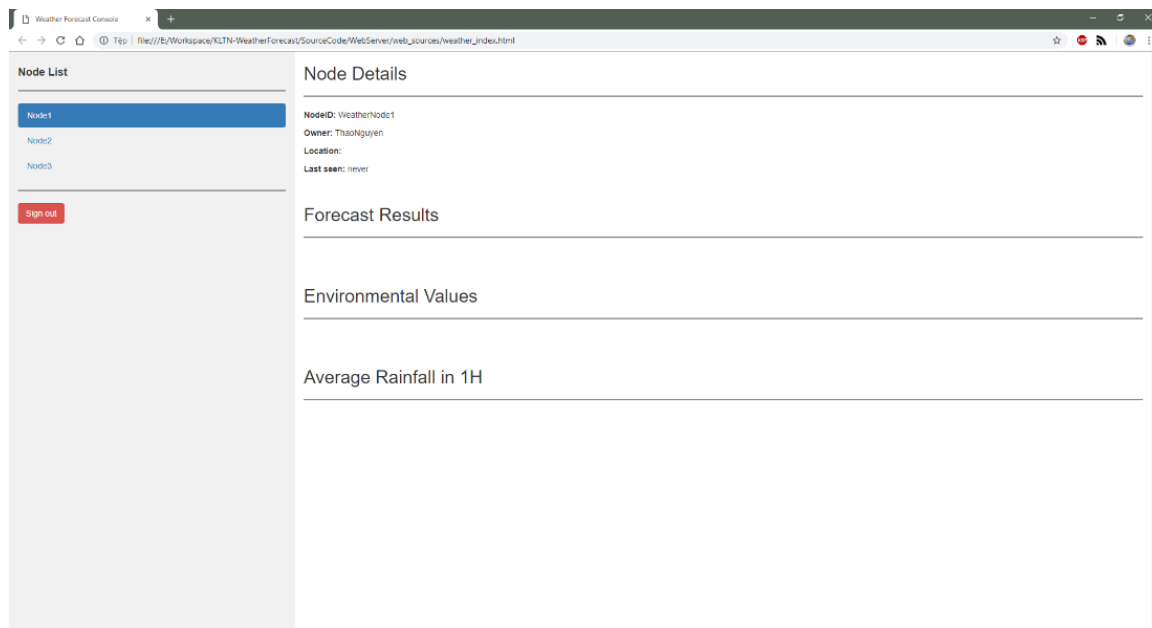
Hình 24: Tổng quan hoạt động của khối quản lý mô hình máy học



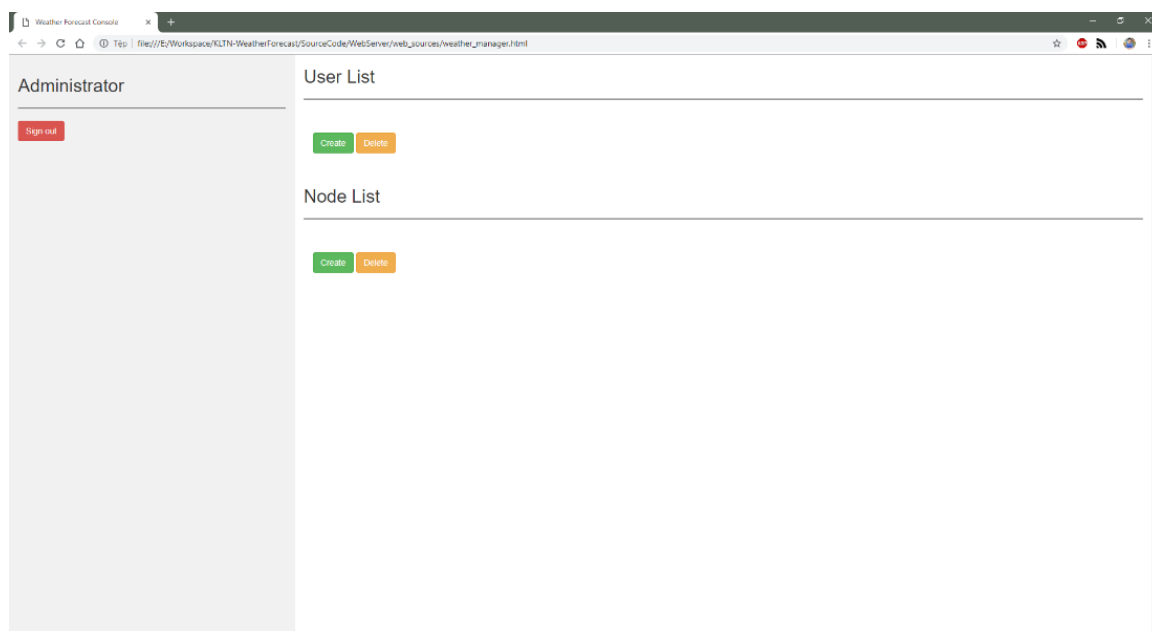
Hình 25: Mô hình máy học được sử dụng trong đề tài



Hình 26: Giao diện đăng nhập trang web điều khiển



Hình 27: Giao diện quan trắc thời tiết trang web điều khiển



Hình 28: Giao diện quản lý trang web điều khiển



Hình 29: Thiết bị quan trắc hoàn chỉnh

Cơ quan Chủ trì
(ký, họ và tên, đóng dấu)

Chủ nhiệm đề tài
(ký, họ và tên)

Nguyễn Mạnh Thảo