

ỨNG DỤNG MẠNG NEURON ĐỂ DỰ BÁO MƯA KẾT HỢP MẠNG CẢM BIẾN KHÔNG DÂY

APPLYING NEURAL NETWORK FOR RAIN FORECAST WITH WIRELESS SENSOR NETWORK

Trần Kim Toại, Dương Cao Trọng Nhân, Võ Minh Huân
Trường Đại học Sư phạm Kỹ thuật TP.HCM, Việt Nam

Ngày toà soạn nhận bài 12/5/2021, ngày phản biện đánh giá 31/5/2021, ngày chấp nhận đăng 28/6/2021.

TÓM TẮT

Trong thời gian gần đây, các thuật toán máy học được ứng dụng rộng rãi cho việc giải quyết các bài toán dự đoán phức tạp trong đó có dự báo thời tiết. Với những ưu điểm vượt trội như tính linh hoạt, độ chính xác cao, đa dạng về ứng dụng, xử lý dữ liệu có tính biến động theo thời gian, các thuật toán máy học mang tính khách quan và đáp ứng được nhiều yêu cầu của thực tiễn hơn so với các phương pháp trước đây. Mục đích nghiên cứu mà tác giả muốn hướng đến là thiết kế hệ thống dự báo mưa dựa trên mạng nơron nhân tạo kết hợp với mạng cảm biến không dây. Mạng nơron nhân tạo sẽ xử lý các thông số môi trường được thu thập từ mạng cảm biến để đưa ra dự báo về sự kiện mưa. Mô hình mạng nơron sẽ được xây dựng và lựa chọn các thông số phù hợp bằng các sai số dự đoán. Hiệu suất của hệ thống dự báo thời tiết với mô hình mạng nơron đã xây dựng sẽ được kiểm chứng thông qua quá trình thực nghiệm với lượng dữ liệu được thu thập từ thực tế. Hệ thống có thể cập nhật liên tục các thông số môi trường tại nhiều địa điểm. Cơ sở dữ liệu sẽ luôn được cập nhật liên tục với các thông số được thu thập theo thời gian thực và đa dạng vì dữ liệu sẽ được thu thập tại nhiều địa điểm lắp đặt các nút cảm biến trong một khu vực lớn triển khai mạng cảm biến không dây để tăng thêm độ tin cậy cho kết quả dự báo. Đồng thời, cơ sở dữ liệu cũng sẽ dựa trên những thông số được lấy từ các website lưu trữ lịch sử thời tiết làm cơ sở cho tập huấn luyện của mô hình hệ thống. Ngoài ra, kết quả dự báo của hệ thống sẽ sự kiện thời tiết sắp tới theo mô hình phân loại (mưa hay không mưa) thay vì các thông số thời tiết, như vậy sẽ dễ dàng hơn cho người dùng hoặc các hệ thống tự động vì nếu kết quả dự báo là các thông số thời tiết thì ta phải phân tích các thông số dự đoán này sau đó mới có thể gửi đến người dùng. Với kết quả dự báo là sự kiện mưa mà tác giả đang hướng đến, từ phía người dùng hay các hệ thống tự động có thể đưa ra quyết định thực thi hay không các tác vụ như tưới tiêu, phun sương, ... để đảm bảo nông sản trong các ứng dụng nông nghiệp chính.

Từ khóa: Mạng thần kinh nhân tạo; Mạng cảm biến không dây; Máy học; Dự báo thời tiết; nông nghiệp thông minh.

ABSTRACT

In recent times, machine learning algorithms are widely applied to solve complex nonlinear problems including weather forecasting. With outstanding advantages such as flexibility, high accuracy, variety of applications, data processing with time fluctuations, machine learning algorithms are objective and meet many practical requirements than the previous methods. The research aims to design a rain forecasting system based on artificial neural networks combined with wireless sensor networks. The neural network processes the environmental parameters collected from the sensor network to make the rain event forecast. A neural network model will be built and selected suitable parameters based on prediction errors. Predictive error is the difference between the real value and the forecasted value to assess the quality or suitability of the forecasting model. The performance of the weather forecasting

system with the built neural network model will be verified through an experimental process with the amount of data collected from reality. The system can continuously update the environmental parameters in many locations. The database will always be constantly updated with various real-time and diverse parameters as data will be collected at multiple sensor nodes installed in a large network deployment area to added reliability for forecast results. At the same time, the database will also be based on parameters taken from weather history storage websites as the basis for the training set of the system model. In addition, the forecast results of the system will be of upcoming weather events according to the classification model (rain or no rain) instead of weather parameters, so it will be easier for users or systems to make a decision because if the forecast results are weather parameters, we have to analyze these predictive parameters before sending them to the user. From the user side, the automated systems can make decisions based on forecasting the rain event whether to execute tasks such as irrigation or misting and assurance of agro-products in agricultural applications.

Keywords: *Neural Network; Wireless Sensor Network; Machine Learning; Weather Forecasting; smart agriculture.*

1. TỔNG QUAN

Ngày nay, sản xuất nông nghiệp ở Việt Nam vẫn phụ thuộc rất nhiều vào thời tiết. Sự thay đổi thất thường của các yếu tố thời tiết sẽ ảnh hưởng trực tiếp đến hoạt động sản xuất nông nghiệp, nhất là trồng trọt. Vậy nên nếu những thay đổi thời tiết có thể được dự đoán với độ chính xác đáng tin cậy thì chúng ta có thể sử dụng những dự đoán này để giúp người nông dân cũng như các hệ thống nông nghiệp tự động đưa ra quyết định điều chỉnh phù hợp giúp cho sản phẩm nông nghiệp luôn ở trong môi trường tốt nhất để phát triển, cũng như tránh được ảnh hưởng hoặc giảm thiểu thiệt hại từ thời tiết xấu [1].

Mặc dù khoa học dự báo thời tiết ra đời đã ra đời từ rất lâu nhưng hiện nay chất lượng công tác dự báo thời tiết vẫn đang không ngừng được các nhà khoa học nghiên cứu và cải thiện. Có rất nhiều phương pháp đang được triển khai hiện nay. Các thuật toán máy học đang được sử dụng rộng rãi để giải quyết các bài toán dự đoán ở khắp nơi trên thế giới. Được biết đến và ứng dụng nhiều phải kể đến những thuật toán như Neural Network, Support Vector Machine (SVM), Naive Bayes, Linear Regression v.v. [2]. Trong đó thuật toán Neural Network đang được sử dụng rất phổ biến trong dự đoán ô nhiễm môi trường, dự đoán thời tiết, dự đoán môi trường gây hại cho cây trồng, v.v. [3-7].

Có rất nhiều nghiên cứu đã khai thác về vấn đề này và cho chúng ta thấy được sự đa dạng về ứng dụng của các thuật toán để giải quyết các bài toán trong thực tế.

Ref. [8] cho chúng ta thấy được sự hiệu quả của mạng cảm biến và máy học để xử lý những thông tin, thực hiện dự đoán. Ngoài ra, những nghiên cứu sau đây cũng đã đề cập rất nhiều đến sự kết hợp của mạng cảm biến không dây và các mô hình máy học [9-12].

Bên cạnh đó, những nghiên cứu về dự đoán được sử dụng rất nhiều trong vấn đề thiên tai, lũ lụt. Có thể kể đến một vài đề tài đã sử dụng mạng cảm biến kết hợp với mô hình Neural Network để giải quyết bài toán dự báo lũ lụt, thiên tai [13-17]. Ngoài những ứng dụng về dự báo lũ lụt, dự báo những thông số thời tiết hằng ngày trong thời gian gần với mô hình mạng Neural Network cũng đã được áp dụng từ rất lâu như Ref [18-19]. Ngoài những ứng dụng về dự báo, các thuật toán máy học còn có thể sử dụng để lọc dữ liệu và giảm thiểu dữ liệu lưu trữ không cần thiết [20].

Thuật toán máy học đang được ứng dụng để xử lý rất nhiều vấn đề trong thực tế với nhiều lĩnh vực khác nhau. Bên cạnh đó, các nghiên cứu trên về dự báo đa phần đều sử dụng cơ sở dữ liệu được lấy từ các website lưu trữ lịch sử thời tiết từ nhiều năm về trước và kết quả dự báo từ hệ thống mà các nghiên

cứu này xây dựng là các thông số thời tiết (nhiệt độ, độ ẩm, tốc độ gió, số giờ nắng ...). Xây dựng hệ thống có thể cập nhật liên tục các thông số môi trường tại nhiều địa điểm là cần thiết. Cơ sở dữ liệu sẽ luôn được cập nhật liên tục với các thông số được thu thập theo thời gian thực và đa dạng vì dữ liệu sẽ được thu thập tại nhiều địa điểm lắp đặt các nút cảm biến trong một khu vực lớn triển khai mạng cảm biến không dây. Ngoài ra, kết quả dự báo của hệ thống sẽ sự kiện thời tiết sắp tới theo mô hình phân loại (mưa hay không mưa) thay vì các thông số thời tiết. Như vậy sẽ dễ dàng hơn cho người dùng hoặc các hệ thống tự động vì nếu kết quả dự báo là các thông số thời tiết thì hệ thống phải phân tích các thông số dự đoán này sau đó mới có thể gửi đến người dùng. Với kết quả dự báo là sự kiện mưa mà nghiên cứu đang hướng đến, từ phía người dùng hay các hệ thống tự động có thể đưa ra quyết định thực thi hay không các tác vụ như tưới tiêu, phun sương, ... để đảm bảo nông sản trong các ứng dụng nông nghiệp.

Hệ thống dự báo mưa mà tác giả đề xuất bao gồm một thiết bị có thể cập nhật dữ liệu môi trường theo thời gian thực và một thuật toán máy học để dự đoán sự kiện mưa trong tương lai. Các dữ liệu môi trường bao gồm nhiệt độ, độ ẩm sẽ được thu thập từ xa sau đó xử lý trên máy tính. Mô hình máy học Neural Network cùng với thuật toán Back Propagation sẽ được triển khai để xử lý dữ liệu môi trường, từ đó sẽ đưa ra dự đoán sự kiện mưa vào ngày tiếp theo. Với sự kiện đã được dự đoán sẽ là cơ sở để đưa ra những quyết định tức thời một cách thuận tiện trong nhiều lĩnh vực như nông nghiệp (điều khiển tưới tiêu, bón phân, điều chỉnh nhiệt độ), công nghiệp (điều khiển động cơ, vận hành hệ thống làm mát), v.v. Bên cạnh đó, những đánh giá về hiệu quả khi áp dụng mô hình Neural Network vào hệ thống dự báo và hiệu năng của mô hình này so với các mô hình thuật toán dự đoán phổ biến khác như: SVM, Logistic Regression, Naive Bayes, v.v. cũng sẽ được thể hiện.

Nghiên cứu thực hiện một hệ thống dự báo thời tiết với độ chính xác cao, cung cấp một công cụ hữu hiệu cho nông dân và các hệ thống nông nghiệp tự động để giảm thiểu các tác động tiêu cực từ thời tiết với môi trường phát triển sản phẩm nông nghiệp.

2. CƠ SỞ LÝ THUYẾT

Hệ thống thu thập và phân tích dữ liệu bao gồm mạng cảm biến không dây dùng để thu thập dữ liệu sử dụng chuẩn truyền dữ liệu không dây Zigbee được đề cập mục 2.1 và 2.2. Mạng thần kinh nhân tạo xử lý dữ liệu và đưa ra quyết định dự báo được phân tích mục 2.3.

2.1 Mạng cảm biến không dây

Mạng cảm biến không dây là hệ thống các nút cảm biến kết nối với nhau bằng sóng vô tuyến và từ đó hình thành một mạng lưới liên kết, trong đó các nút này thường là các thiết bị nhỏ gọn, giá thành thấp và có khả năng thu thập các thông số môi trường xung quanh khu vực lắp đặt. Các điểm cảm biến này được lắp đặt số lượng lớn và phân bố một cách không có hệ thống trên một diện tích rộng, có thời gian hoạt động lâu dài và sử dụng nguồn năng lượng hạn chế [21].

Một nút cảm biến được cấu tạo bởi 4 thành phần cơ bản sau : bộ xử lý, bộ cảm biến, bộ thu phát không dây và nguồn cung cấp. Tùy theo mục đích ứng dụng, nút cảm biến còn có thể có các thành phần bổ sung như hệ thống định vị và thiết bị di động [22].

2.2 Công nghệ truyền thông không dây Zigbee

Công nghệ ZigBee được xây dựng dựa trên tiêu chuẩn 802.15.4 của tổ chức IEEE. Tiêu chuẩn 802.15.4 này sử dụng tín hiệu radio có tần số ngắn [23]. Mục tiêu của công nghệ ZigBee là nhắm tới việc truyền tin với mức tiêu hao năng lượng nhỏ và công suất thấp cho những thiết bị chỉ có thời gian hoạt động từ vài tháng đến vài năm mà không yêu cầu cao về tốc độ truyền tin [24].

2.3 Mạng thần kinh nhân tạo

Mạng thần kinh nhân tạo (Artificial Neural Network) là mô hình xử lý thông tin

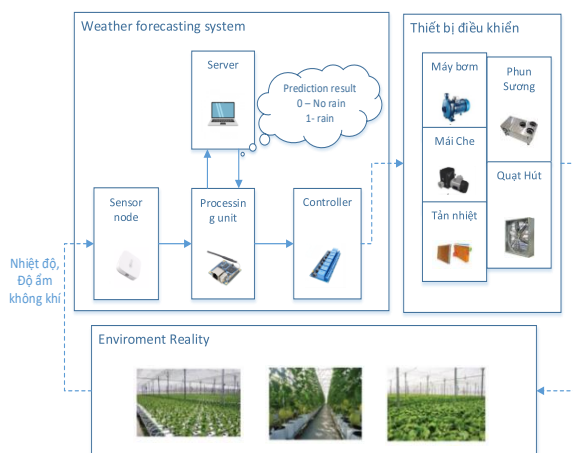
được mô phỏng dựa trên hoạt động của hệ thống thần kinh của sinh vật, bao gồm số lượng lớn các Neuron được gắn kết để xử lý thông tin và kết nối với nhau thông qua các liên kết (gọi là trọng số liên kết), làm việc như một thể thống nhất để giải quyết một vấn đề cụ thể nào đó [25-26].

Mạng nơron nhân tạo là hệ thống bao gồm nhiều nơron nhân tạo kết hợp với nhau. Mỗi nơron có các đặc tính đầu vào, đầu ra và thực hiện một số chức năng tính toán cục bộ. Mạng nơron có khả năng học, khái quát hóa tập dữ liệu học thông qua việc gán và hiệu chỉnh các trọng số liên kết [27-29].

3. THIẾT KẾ HỆ THỐNG

3.1 Tổng quan hệ thống

Các dữ liệu thời tiết để cập nhật cho tập dữ liệu đầu vào của mô hình được đo lường trực tiếp bằng các cảm biến nhiệt độ, độ ẩm. Tuy nhiên, ở các vị trí khác nhau thì thông số thời tiết sẽ có sự thay đổi nhất định, vì vậy nhiều nút cảm biến



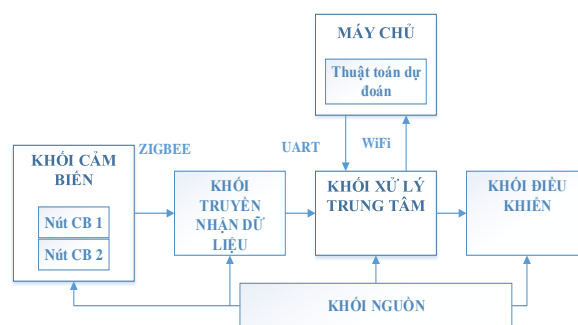
Hình 1. Sơ đồ hoạt động của hệ thống được đặt ở nhiều vị trí khác nhau.

Hình 1 trình bày sơ đồ hoạt động của hệ thống thu thập và xử lý dữ liệu. Các thông số nhiệt độ, độ ẩm của từng nút cảm biến sẽ được gửi đến bộ xử lý trung tâm, việc truyền nhận dữ liệu sẽ thông qua thiết bị phần cứng có thể giao tiếp theo chuẩn không dây với khoảng cách phù hợp. Bộ xử lý trung tâm sẽ đưa dữ liệu đến máy chủ – nơi thực hiện quá trình huấn luyện của thuật toán lan truyền ngược. Dữ liệu trên máy chủ sẽ là đầu vào của

mô hình mạng nơron truyền thẳng nhiều lớp. Sau khi huấn luyện, kết quả đầu ra của mô hình chính là kết quả dự báo sự kiện mưa.

3.2 Sơ đồ phần cứng hệ thống và lưu đồ hoạt động

Sơ đồ chi tiết hệ thống phần cứng sẽ gồm các khối được thể hiện ở hình sau:



Hình 2. Sơ đồ phần cứng hệ thống

Chức năng thành phần của hệ thống được trình bày trong hình 2:

Khối nguồn có chức năng cung cấp nguồn cho tất cả các mạch trong hệ thống.

Khối xử lý trung tâm sẽ liên lạc giữa khối truyền nhận dữ liệu thông qua chuẩn truyền thông UART và với máy chủ thông qua WiFi. Tập hợp các giá trị thu thập được từ cảm biến sau đó lần lượt gửi về máy chủ. Đồng thời, khối trung tâm cũng sẽ gửi tín hiệu đến khối điều khiển sau khi máy chủ đưa ra dự đoán.

Khối truyền nhận dữ liệu là một trạm trung gian để nhận dữ liệu từ các nút cảm biến cũng như truyền dữ liệu đến khối xử lý trung tâm.

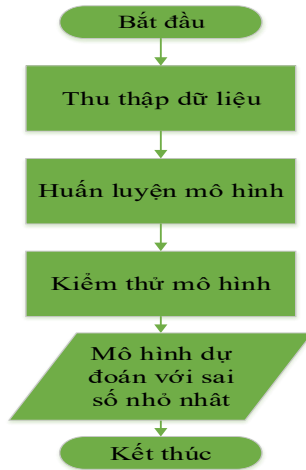
Nút cảm biến có chức năng thu thập các giá trị nhiệt độ, độ ẩm không khí ở nhiều khu vực khác nhau.

Máy chủ xử lý các dữ liệu nhận từ khối xử lý trung tâm và đưa tới đầu vào của mô hình thuật toán dự đoán và huấn luyện mô hình dựa trên những dữ liệu được cung cấp. Đồng thời đưa ra kết quả dự đoán sau khi hoàn thành huấn luyện mô hình.

Khối điều khiển sẽ nhận tín hiệu từ khối xử lý trung tâm để thực thi đóng mở các thiết

bị sau khi thuật toán đưa ra kết quả dự báo về sự kiện mưa sắp tới.

Lưu đồ của hệ thống sẽ gồm các bước được thể hiện như hình 3.



Hình 3. Lưu đồ hệ thống

3.3 Xây dựng mô hình dự báo

Các bước xây dựng một mạng Neuron bao gồm:

Bước 1: Chuẩn hóa dữ liệu đầu vào và phân chia tập dữ liệu.

Bước 2: Thiết kế cấu trúc mạng

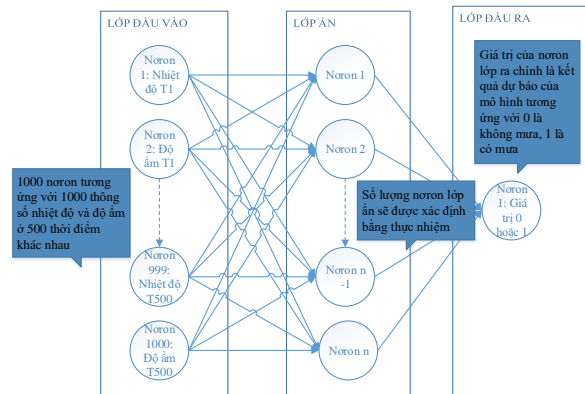
Bước 3: Khởi tạo giá trị ngẫu nhiên cho trọng số.

Bước 4: Thực hiện giai đoạn lan truyền tiến đi qua toàn bộ các lớp của mạng.

Bước 5: Tính sai số trên bộ học ở đầu ra và quyết định xem có học tiếp hay không, nếu học tiếp thì chuyển đến bước 5 nếu không thì dừng vòng lặp.

Bước 6: Tính sai số trên từng nút của mỗi lớp để cập nhật trọng số và tiếp tục lặp lại bước 3.

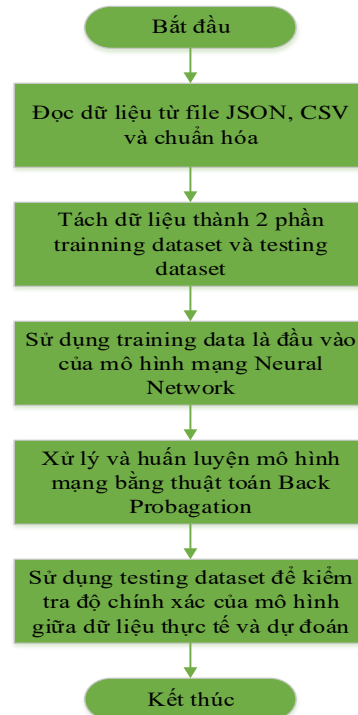
Hình 4 dưới đây là cấu trúc của mạng neuron truyền thẳng với 3 lớp (1 lớp vào, 1 lớp ẩn, 1 lớp ra) sau khi đã lựa chọn:



Hình 4. Mô hình Neural Network thực nghiệm

3.4 Huấn luyện mô hình mạng

Lưu đồ huấn luyện mô hình mạng neuron được trình bày trong hình 5.



Hình 5. Lưu đồ huấn luyện Neural Network

Tập dữ liệu ban đầu được chia thành 2 tập dữ liệu thành phần, bộ dữ liệu thứ nhất dùng làm dữ liệu đầu vào cho bài toán dự báo, bộ dữ liệu thứ hai dùng để kiểm tra kết quả của phương pháp dự báo. Trong bài toán dự báo sử dụng mạng neuron nhân tạo, bộ số liệu thứ nhất sẽ được đưa vào mạng để huấn luyện. Sau mỗi lần lặp của quá trình huấn luyện, mạng sẽ lưu lại ma trận trọng số và đối chiếu sai số của mạng với sai số cho phép ban đầu.

Kết quả của quá trình huấn luyện là mô hình mạng đã được huấn luyện với tập dữ liệu chính là ma trận trọng số có sai số nhỏ nhất.

4. THỰC NGHIỆM VÀ ĐÁNH GIÁ

4.1 Tiền xử lý dữ liệu

Dữ liệu cần thiết để làm đầu vào cho mô hình dự đoán là nhiệt độ: đơn vị là °C và độ ẩm không khí: đơn vị là %. Để kết quả dự đoán có độ tin cậy cao tác giả sẽ sử dụng hai nguồn dữ liệu để xây dựng cơ sở dữ liệu cho mô hình mạng huấn luyện. Bên cạnh dữ liệu từ các website lưu trữ lịch sử thời tiết. Cơ sở dữ liệu cũng sẽ được cập nhật liên tục với các dữ liệu được thu thập theo thời gian thực từ các nút cảm biến được triển khai mô hình mạng cảm biến không dây.

Hai nút cảm biến sẽ được đặt ở 2 khu vực cách xa nhau khoảng 500m và thu thập dữ liệu liên tục với thời gian cách nhau là 5 phút giữa các lần đo, sau đó gửi đến khối xử lý trung tâm, dữ liệu tại đây sẽ được lưu trữ lên máy chủ dưới dạng file json như được trình bày trong hình 6 và sẵn sàng để đưa đến đầu vào của mô hình huấn luyện dự báo.

```
[{"Time": "18:19:14 27-04-2020", "Name": "SMART-02028327F5_temp", "Temp": "25"}, {"Time": "18:19:14 27-04-2020", "Name": "SMART-02025CE097_humi", "Humi": "60"}, {"Time": "18:19:14 27-04-2020", "Name": "SMART-02028327F5_humi", "Humi": "67"}, {"Time": "18:19:14 27-04-2020", "Name": "SMART-02025CE097_temp", "Temp": "26"}, {"Time": "18:24:14 27-04-2020", "Name": "SMART-02028327F5_temp", "Temp": "25"}, {"Time": "18:24:14 27-04-2020", "Name": "SMART-02025CE097_humi", "Humi": "60"}, {"Time": "18:24:14 27-04-2020", "Name": "SMART-02028327F5_humi", "Humi": "67"}, {"Time": "18:24:14 27-04-2020", "Name": "SMART-02025CE097_temp", "Temp": "26"}, {"Time": "18:29:14 27-04-2020", "Name": "SMART-02028327F5_temp", "Temp": "25"}, {"Time": "18:29:14 27-04-2020", "Name": "SMART-02025CE097_humi", "Humi": "60"}, {"Time": "18:29:14 27-04-2020", "Name": "SMART-02028327F5_humi", "Humi": "67"}, {"Time": "18:29:14 27-04-2020", "Name": "SMART-02025CE097_temp", "Temp": "26"}, {"Time": "18:34:14 27-04-2020", "Name": "SMART-02028327F5_temp", "Temp": "25"}, {"Time": "18:34:14 27-04-2020", "Name": "SMART-02025CE097_humi", "Humi": "59"}, {"Time": "18:34:14 27-04-2020", "Name": "SMART-02028327F5_humi", "Humi": "67"}, {"Time": "18:34:14 27-04-2020", "Name": "SMART-02025CE097_temp", "Temp": "26"}, {"Time": "18:39:14 27-04-2020", "Name": "SMART-02028327F5_temp", "Temp": "25"}, {"Time": "18:39:14 27-04-2020", "Name": "SMART-02025CE097_humi", "Humi": "59"}, {"Time": "18:39:14 27-04-2020", "Name": "SMART-02028327F5_humi", "Humi": "67"}, {"Time": "18:39:14 27-04-2020", "Name": "SMART-02025CE097_temp", "Temp": "26"}, {"Time": "18:44:14 27-04-2020", "Name": "SMART-02028327F5_temp", "Temp": "25"}, {"Time": "18:44:14 27-04-2020", "Name": "SMART-02025CE097_humi", "Humi": "58"}, {"Time": "18:44:14 27-04-2020", "Name": "SMART-02028327F5_humi", "Humi": "66"}, {"Time": "18:44:14 27-04-2020", "Name": "SMART-02025CE097_temp", "Temp": "26"}]
```

Hình 6. Thông số thời tiết json file

Ngoài ra, dữ liệu nhiệt độ và độ ẩm không khí trên website dự báo thời tiết tại thành phố Hồ Chí Minh cũng sẽ là đầu vào của mô hình huấn luyện dự báo. Các dữ liệu này được thu thập với thời gian cách nhau là 1 tiếng và được lưu dưới dạng file csv trong hình 7.

	A	B	C
1	Địa điểm	Ho Chi Minh City	Ho Chi Minh City
2	Thời gian lấy mẫu (mm/dd/yy hour)	Nhiệt độ (°C)	Độ ẩm không khí (%)
3	5/24/20 0:00	28.453207 °C	76%
4	5/24/20 1:00	27.983208 °C	78%
5	5/24/20 2:00	27.563208 °C	80%
6	5/24/20 3:00	27.183207 °C	81%
7	5/24/20 4:00	26.893208 °C	82%
8	5/24/20 5:00	26.793207 °C	83%
9	5/24/20 6:00	27.013206 °C	82%
10	5/24/20 7:00	28.593206 °C	75%
11	5/24/20 8:00	30.423206 °C	67%
12	5/24/20 9:00	31.723207 °C	62%
13	5/24/20 10:00	31.013206 °C	64%
14	5/24/20 11:00	30.993208 °C	63%
15	5/24/20 12:00	31.043207 °C	62%
16	5/24/20 13:00	31.013206 °C	63%
17	5/24/20 14:00	30.963207 °C	63%
18	5/24/20 15:00	30.803207 °C	64%

Hình 7. Thông số thời tiết csv file

Các giá trị của hai thông số này tương đối chênh lệch nhau. Sự chênh lệch này yêu cầu phải chuẩn hóa dữ liệu đầu vào để các thông số đầu vào của mạng có vai trò như nhau và chúng ảnh hưởng đến kết quả đầu ra ở mức độ nhất định.

4.2 Lựa chọn dữ liệu

Các dữ liệu nhiệt độ, độ ẩm sẽ được chia thành 2 tập dữ liệu bao gồm Training dataset và Testing dataset như được biểu diễn trong Bảng 1. Training dataset dùng để làm đầu vào để thực thi thuật toán huấn luyện mô hình mạng. Testing dataset dùng để đánh giá độ chính xác của mô hình mạng sau khi huấn luyện

Bảng 1. Bảng phân chia dữ liệu huấn luyện và kiểm thử

Dataset	Training dataset	Testing dataset
Website	5000 data (70%)	2000 data (30%)
Sensor	65000 data (70%)	28000 data (30%)

Hai tập dữ liệu này sẽ tập hợp các dữ liệu được lấy từ 2 nguồn sau:

Dữ liệu nhiệt độ, độ ẩm từ cơ sở dữ liệu trên website dự báo thời tiết tại thành phố Hồ Chí Minh (từ 0h 1/4/2020 đến 0h 1/9/2020) và lưu trữ ở máy chủ dưới dạng file csv.

Dữ liệu nhiệt độ, độ ẩm từ các nút cảm biến được thu thập tại các khu vực đặt nút

cảm biến (từ 0h 20/4/2020 đến 0h 25/7/2020) và lưu trữ trên máy chủ dưới dạng file json.

4.3 Thực nghiệm thuật toán

Tập dữ liệu training dataset ở bảng 1 sẽ được chia thành các mảng dữ liệu. Mỗi tập sẽ có 1000 dữ liệu nhiệt độ, độ ẩm ở 500 thời điểm khác nhau và sẽ được phân chia vào 2 nhóm trên. Các tập dữ liệu này sẽ lần lượt được đưa vào để huấn luyện cho mạng bao gồm 70 tập dữ liệu để tăng khả năng học của mạng với nhiều trường hợp thời tiết khác nhau.

Đầu ra mong muốn của mô hình này là 0 (không mưa) và 1 (có mưa) sẽ được so sánh với đầu ra huấn luyện của mô hình sau mỗi vòng lặp để tính toán sai số của mạng ở mỗi lần lặp. Sai số hàm lỗi và đầu ra huấn luyện sẽ được cập nhật sau mỗi vòng lặp huấn luyện khi ma trận trọng số đã được thay đổi ở lần lặp trước. Ma trận trọng số sẽ được khởi tạo ngẫu nhiên ở lần lặp đầu tiên.

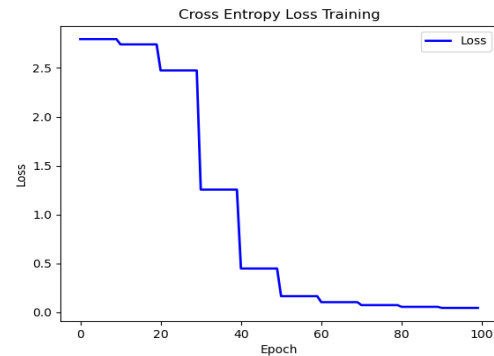
Bảng 2. Huấn luyện mô hình mạng với số nơron lớp ẩn khác nhau

Số liệu của mô hình mạng	Số liệu dự báo ứng với số lượng nơron lớp ẩn				
	10	50	100	200	500
Đầu ra mô hình (rain node)	0.98866309	0.98887655	0.98885199	0.98804823	0.98769783
Đầu ra mô hình (no rain node)	0.01319132	0.01265871	0.01215356	0.0129603	0.01350966
Sai số hàm mất mát	0.048	0.047	0.046	0.049	0.05
Sai số MSE	0.000102	0.000096	0.000093	0.000105	0.000112
Sai số MAE	0.0101	0.0098	0.0096	0.0102	0.0106
Sai số RMSE	0.0101	0.0098	0.0097	0.0103	0.0106

Thực nghiệm trong bảng 2 cho thấy, số nơron lớp ẩn là 100 nơron dự báo tốt nhất với sai số thấp nhất trên cả 3 phương pháp đánh giá sai số MSE, RMSE và MAE. Trong 3 phương pháp trên thì phương pháp tính sai số MSE cho sai số thấp nhất. Từ bảng 2, ta thấy nếu số lượng nơron lớp ẩn quá ít hoặc quá nhiều thì mạng sẽ cho độ chính xác không cao. Đặc biệt số lượng nơron lớp ẩn quá lớn thì sai số giảm không đáng kể và quá trình

học sẽ diễn ra lâu hơn dẫn đến việc tiêu tốn tài nguyên phần cứng.

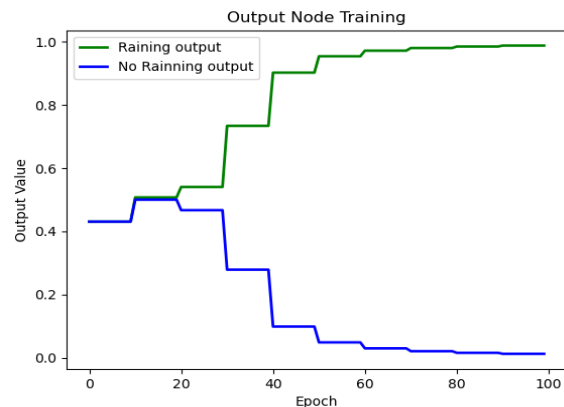
Với kết quả trên, ta thực hiện huấn luyện mô hình 100 nơron ẩn với 100 vòng lặp để quan sát được quá trình huấn luyện của mạng. Đồng thời theo dõi kết quả ở từng lần lặp để đánh giá quá trình học.



Hình 8. Hàm mất mát Cross Entropy với 100 vòng lặp

Với kết quả ở hình 8, ta có thể thấy hàm mất mát giảm từ 2.7287201 xuống 0.05124195 sau 100 lần lặp của quá trình huấn luyện. Hàm mất mát của mô hình đang dần tiến đến cực tiểu lý tưởng là 0 qua các lần lặp huấn luyện.

Với đầu ra dự đoán lý tưởng của nút ra dự đoán mưa là 1 và nút ra dự đoán không mưa là 0. Sau 100 lần lặp của quá trình huấn luyện, ta thấy đầu ra của mạng giảm từ 0.518 xuống 0.013 với nút đầu ra dự đoán không mưa và tăng từ 0.519 lên 0.987 với nút đầu ra dự đoán có mưa. Với kết quả ở hình 9, ta có thể thấy các nút đầu ra của mạng đang tiến dần đến dự đoán lý tưởng sau mỗi vòng lặp huấn luyện.



Hình 9. Đầu ra mô hình với 100 vòng lặp

4.4 Kiểm thử thuật toán

Tập dữ liệu testing dataset sẽ được chia thành 30 tập test với mỗi tập test là 1000 dữ liệu nhiệt độ và độ ẩm ở 500 thời điểm khác nhau. Ta sẽ thực hiện 30 lần kiểm thử tương ứng với 30 tập test này bằng mô hình mạng đã được huấn luyện để đánh giá khả năng dự đoán của mạng sau khi đã huấn luyện.

Bảng 3. Kiểm thử với tập dữ liệu testing dataset

Confusion matrix	Predict: No	Predict: Yes
Actual: No	13	0
Actual: Yes	3	14

Với bảng 3, ta có 27 lần dự đoán đúng trên tổng số 30 lần kiểm thử. Trong đó, 13 đại diện cho số lần dự đoán đúng rằng mưa sẽ không xảy ra và 14 là số dự đoán đúng rằng mưa sẽ xảy ra. Ngược lại, ta có 3 dự đoán sai trên tổng số 30 lần kiểm thử. Trong đó, 3 là số lần dự đoán không mưa nhưng thực tế là mưa đã xảy ra và 0 là số lần dự đoán có mưa nhưng thực tế là mưa đã không xảy ra.

4.5 So sánh thuật toán

Để chứng minh tính hiệu quả của thuật toán lan truyền ngược với mô hình mạng nơron nhiều lớp, tác giả tiến hành huấn luyện và kiểm thử so sánh với các thuật toán khác như Naive Bayes, Support Vector Machine, K – Nearest Neighbor, Decision Tree, Random Forest, Logistic Regression [2] với cùng tập dữ liệu huấn luyện. Kết quả đánh giá so sánh được trình bày trong Bảng 4 với so sánh thời gian huấn luyện và Bảng 5 với phần trăm độ chính xác.

Bảng 4. Thời gian huấn luyện của các thuật toán

Algorithm	Training time (s)
Neural Network	0.57
Naive Bayes	0.007

Support Vector Machine	0.005
K – Nearest Neighbor	0.003
Decision Tree	0.008
Random Forest	0.028
Logistic Regression	0.048

So sánh kết quả độ chính xác của các thuật toán ở bảng 5 với thời gian huấn luyện ở bảng 4, ta có thể thấy rằng các thuật toán có thời gian huấn luyện dài sẽ cho kết quả dự đoán tốt hơn so với các mô hình thuật toán có thời gian huấn luyện ngắn. Mô hình thuật toán Neural Network với thời gian huấn luyện là 0.57 giây tương ứng với độ chính xác là 90%. Kế tiếp là Logistic Regression với 0.048 tương ứng với 86.6%. Các thuật toán Naive Bayes, SVM, KNN, Decision Tree có tỉ lệ thấp hơn, dao động từ 76.6% đến 83.3% nhưng có thời gian huấn luyện ngắn hơn so với 2 mô hình thuật toán trước, dao động từ 0.003 – 0.008 giây. Điều này cho ta thấy rằng thời gian huấn luyện mô hình sẽ tỷ lệ nghịch với độ chính xác của mô hình thuật toán.

Bảng 5. Phần trăm dự đoán của các thuật toán

Algorithm	Accuracy (%)
Neural Network	90
Naive Bayes	76.6
Support Vector Machine	80
K – Nearest Neighbor	76.6
Decision Tree	83.3
Random Forest	60
Logistic Regression	86.6

Tuy nhiên, mô hình thuật toán Random Forest có thời gian huấn luyện lâu hơn các thuật toán Naive Bayes, SVM, KNN, Decision Tree nhưng độ chính xác dự đoán lại khá thấp với 60%. Điều này chứng tỏ Random Forest chưa huấn luyện hiệu quả với các thông số đã chọn hoặc mô hình này chưa thích hợp để xử lý các thông số dữ liệu về thời tiết.

5. KẾT LUẬN

Trong bài báo này, tác giả đã xây dựng một mô hình thu thập dữ liệu ổn định, tiêu tốn ít năng lượng, các điểm thu thập hoạt động ở nhiều khu vực khác nhau. Các thực nghiệm cho thấy, nếu như được huấn luyện tốt trên tập các dữ liệu đầy đủ và hoàn thiện với các tham số được lựa chọn cẩn thận thì kết quả dự báo đạt độ chính xác cao tới 90% với mô hình sử dụng mạng neuron so với các mô hình khác như Logistic Regression với

86.6%. Các thuật toán Naive Bayes, SVM, KNN, Decision Tree có tỉ lệ thấp hơn, dao động từ 76.6% đến 83.3%.

LỜI CẢM ƠN

Bài báo được thực hiện với sự hỗ trợ kinh phí từ dự án số: T2020-48TĐ, bởi Trường Đại học Sư phạm Kỹ thuật TP.HCM, Việt Nam.

TÀI LIỆU THAM KHẢO

- [1] National strategy on climate change. Internet: [Strategies details \(chinhphu.vn\)](http://Strategies%20details%20(chinhphu.vn)), 23/12/2020.
- [2] S. Ray. A Quick Review of Machine Learning Algorithms. 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019, pp. 35-39.
- [3] M. J. Kennedy. Review of applications of artificial neural networks in biotechnology, Proceedings 1993 The First New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems, 1993, pp. 252-254.
- [4] T. Schmidt, H. Rahnama and A. Sadeghian. A review of applications of artificial neural networks in cryptosystems, 2008 World Automation Congress, 2008, pp. 1-6.
- [5] Y. Li and W. Ma. Applications of Artificial Neural Networks in Financial Economics: A Survey, 2010 International Symposium on Computational Intelligence and Design, 2010, pp. 211-214.
- [6] B. Milovanovic, M. Agatonovic, Z. Stankovic, N. Doncov and M. Sarevska. Application of neural networks in spatial signal processing (invited paper), 11th Symposium on Neural Network Applications in Electrical Engineering, 2012, pp. 5-14.
- [7] R. Meenal and A. I. Selvakumar, Review on artificial neural network based solar radiation prediction, 2017 2nd International Conference on Communication and Electronics Systems (ICCES), 2017, pp. 302-305.
- [8] T. Truong, A. Dinh and K. Wahid. An IoT environmental data collection system for fungal detection in crop fields, 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), 2017, pp. 1-4.
- [9] V. F. Silva, D. F. Macedo and J. L. Leoni. Spectrum Decision in Wireless Sensor Networks Employing Machine Learning, 2014 Brazilian Symposium on Computer Networks and Distributed Systems, 2014, pp. 386-393.
- [10] H. Ahmadi and R. Bouallegue. Comparative study of learning-based localization algorithms for Wireless Sensor Networks: Support Vector regression, Neural Network and Naïve Bayes, 2015 International Wireless Communications and Mobile Computing Conference (IWCMC), 2015, pp. 1554-1558.
- [11] S. K. and V. Vaidehi. Clustering and Data Aggregation in Wireless Sensor Networks Using Machine Learning Algorithms, 2018 International Conference on Recent Trends in Advance Computing (ICRTAC), 2018, pp. 109-115.
- [12] T. Hemalatha, M. V. Ramesh and V. P. Rangan. Effective and Accelerated Forewarning of Landslides Using Wireless Sensor Networks and Machine Learning, in *IEEE Sensors Journal*, vol. 19, no. 21, pp. 9964-9975, 1 Nov.1, 2019.
- [13] J. K. Roy, D. Gupta and S. Goswami. An improved flood warning system using WSN and Artificial Neural Network. 2012 Annual IEEE India Conference, 2012, pp. 770-774.

- [14] I. R. Widiyari, L. E. Nugroho and Widyawan. Deep learning multilayer perceptron (MLP) for flood prediction model using wireless sensor network based hydrology time series data mining. *2017 International Conference on Innovative and Creative Information Technology (ICITech)*, 2017, pp. 1-5.
- [15] S. Madan, P. Kumar, S. Rawat and T. Choudhury. Analysis of Weather Prediction using Machine Learning & Big Data, *2018 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, 2018, pp. 259-264.
- [16] G. Verma, P. Mittal and S. Farheen. Real Time Weather Prediction System Using IOT and Machine Learning, *2020 6th International Conference on Signal Processing and Communication (ICSC)*, 2020, pp. 322-324.
- [17] R. K. Grace and B. Suganya. Machine Learning based Rainfall Prediction, *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2020, pp. 227-229.
- [18] Mohsen Hayati, Zahra Mohebi. Application of Artificial Neural Networks for Temperature Forecasting. *World Academy of Science, Engineering and Technology*, Vol 4, Issue 1, pp. 662-666, 2007.
- [19] A. Molano-Jimenez, A. D. Orjuela-Cañón and W. Acosta-Burbano. Temperature and Relative Humidity Prediction in Swine Livestock Buildings, *2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 2018, pp. 1-4.
- [20] L. P. Dinesh Kumar, S. Shakena Grace, A. Krishnan, V. M. Manikandan, R. Chinraj and M. R. Sumalatha. Data filtering in wireless sensor networks using neural networks for storage in cloud, *2012 International Conference on Recent Trends in Information Technology*, 2012, pp. 202-205.
- [21] H. Sharma and S. Sharma. A review of sensor networks: Technologies and applications, *2014 Recent Advances in Engineering and Computational Sciences (RAECS)*, 2014, pp. 1-4.
- [22] Ian F. Akyildiz and Mehmet Can Vuran. *Wireless Sensor Network*. John Wiley & Sons Ltd, 2010, pp. 10-15.
- [23] Jose A. Gutierrez; Edgar H. Callaway; Raymond L. Barrett. IEEE Std 802.15.4 Technical Overview. *Low-Rate Wireless Personal Area Networks: Enabling Wireless Sensors with IEEE 802.15.4*, 2007.
- [24] C. M. Ramya, M. Shanmugaraj and R. Prabakaran. Study on ZigBee technology. *2011 3rd International Conference on Electronics Computer Technology*, 2011, pp. 297-301.
- [25] A. Sperduti. An overview on supervised neural networks for structure. *Proceedings of International Conference on Neural Networks*, 1997, pp. 2550-2554.
- [26] M. Chen, U. Challita, W. Saad, C. Yin and M. Debbah. Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial. *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3039-3071, 2019.
- [27] A. Menon, S. Singh and H. Parekh. A Review of Stock Market Prediction Using Neural Networks, *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2019, pp. 1-6.
- [28] M. Verhelst and B. Moons, Embedded Deep Neural Network Processing: Algorithmic and Processor Techniques Bring Deep Learning to IoT and Edge Devices. *IEEE Solid-State Circuits Magazine*, vol. 9, no. 4, pp. 55-65, Fall 2017.
- [29] S. C. Sreenivasa, S. K. Agarwal and R. Kumar. Short term wind forecasting using logistic regression driven hypothesis in artificial neural network. *2014 6th IEEE Power India International Conference (PIICON)*, 2014, pp. 1-6.

Tác giả chịu trách nhiệm bài viết:

Võ Minh Huân

Trường Đại học Sư phạm Kỹ thuật TP.HCM
Email: huanvm@hcmute.edu.vn