

Ứng dụng mạng nơ ron nhân tạo trong bài toán ước lượng quãng đường ô tô điện từ thông số kỹ thuật

¹ThS.NCS. Phạm Thị Quỳnh Trang

¹Trường Điện-Điện tử, Đại học Công nghiệp Hà Nội

Ngày nhận bài: 14/11/2025

²ThS. Đinh Trọng Toàn

²Trường Đại học Hải Dương

Ngày chấp nhận đăng: 21/12/2025

³ThS.NCS. Nguyễn Thị Toàn

³Trường Đại học Hải Dương
Email: uhdnguyentoan.edu@gmail.com

Tóm tắt - Dự đoán quãng đường di chuyển của xe điện (Electric Vehicles - EVs) là một yếu tố quan trọng, có ảnh hưởng trực tiếp đến mức độ chấp nhận của người dùng cũng như các chiến lược thiết kế và tiếp thị của nhà sản xuất. Tuy nhiên, các thông số quãng đường chính thức thường chỉ được công bố ở giai đoạn thương mại hóa, trong khi nhu cầu ước lượng chính xác ngay từ giai đoạn thiết kế ngày càng trở nên cấp thiết. Trong bài báo này, chúng tôi đề xuất một mô hình mạng nơ ron nhân tạo kết hợp với khung học máy (Machine Learning - ML) nhằm dự đoán quãng đường di chuyển của xe điện dựa trên các thông số kỹ thuật sẵn có, bao gồm dung lượng pin, hiệu suất năng lượng, mô-men xoắn, kích thước xe và hệ dẫn động. Nhiều mô hình học máy khác nhau được triển khai và so sánh về độ chính xác, bao gồm K-Nearest Neighbors, Decision Tree, Random Forest, Support Vector Machine và mạng nơ ron nhân tạo. Kết quả thực nghiệm cho thấy các phương pháp học máy có khả năng cung cấp ước lượng quãng đường hiệu quả, qua đó bổ sung và hỗ trợ các mô hình dựa trên nguyên lý vật lý truyền thống.

Từ khóa - Xe điện (EVs), học máy, mạng nơ ron nhân tạo (ANN), hiệu quả năng lượng

I. GIỚI THIỆU

Trong thập kỷ qua, xe điện (EV) đã trở thành trọng tâm trong xu hướng phát triển bền vững của ngành công nghiệp ô tô. Một trong những yếu tố quan trọng quyết định mức độ phổ biến của EVs là quãng đường di chuyển sau một lần sạc (range). Đây không chỉ là mối quan tâm hàng đầu của người tiêu dùng mà còn là thước đo cạnh tranh giữa các dòng xe [1].

Các phương pháp truyền thống để dự đoán quãng đường chủ yếu dựa trên mô hình vật lý, tính toán động lực học xe, khí động học và tiêu thụ năng lượng [2]. Tuy nhiên, chúng đòi hỏi dữ liệu chi tiết về động cơ, trọng lượng, điện trở lăn, hệ số cản gió... vốn khó có được ở giai đoạn thiết kế ban đầu. Với sự phát triển của dữ liệu mở và công nghệ học máy, cách tiếp cận dựa trên dữ liệu (data-driven) đang nổi lên như một giải pháp thay thế đầy hứa hẹn.

Bài báo này nhằm xây dựng một mô hình dự đoán quãng đường xe điện dựa trên thông số kỹ thuật công khai. Đóng góp chính gồm:

- Đề xuất một mô hình học máy để ước lượng quãng đường từ các thuộc tính kỹ thuật cơ bản.
- So sánh hiệu quả giữa nhiều thuật toán học máy khác nhau.

- Cung cấp cơ sở dữ liệu thực nghiệm cho thấy ML có thể hỗ trợ quá trình thiết kế và ra quyết định trong ngành công nghiệp xe điện.

II. CÁC NGHIÊN CỨU LIÊN QUAN

2.1. Mô hình dựa trên vật lý

Nhiều nghiên cứu sử dụng các mô hình động lực học và cân bằng năng lượng để ước tính tiêu thụ điện năng. Ví dụ, công trình của Ehsani và cộng sự [2] đưa ra mô hình tính toán mức tiêu thụ năng lượng dựa trên lực cản gió, trọng lượng và đặc tính động cơ. Hu và cộng sự [3] cũng chỉ ra rằng các mô hình này phù hợp cho thiết kế chi tiết, nhưng hạn chế trong dự báo sớm khi chưa có đầy đủ thông số kỹ thuật.

2.2. Mô hình dựa trên dữ liệu hành trình

Một hướng phổ biến là sử dụng dữ liệu thực tế từ cảm biến và hành trình lái xe. Các nghiên cứu này sử dụng dữ liệu hành trình thực tế để học mối quan hệ giữa đặc trưng vận hành (như vận tốc theo tuyến đường) và mức tiêu thụ năng lượng, ứng dụng mạng nơ ron sâu và các mô hình học máy data-driven, cho kết quả vượt trội so với mô hình tuyến tính truyền thống [4][6]. Tuy nhiên, các mô hình này đòi hỏi dữ liệu lái xe chi tiết, khó áp dụng trong giai đoạn thiết kế.

2.3. Mô hình dựa trên dữ liệu thông số kỹ thuật

Một số nghiên cứu gần đây bắt đầu chú ý đến việc dự đoán quãng đường chỉ từ dữ liệu tĩnh (thông số kỹ thuật xe). Zhang và cộng sự [5] đã tổng hợp các phương pháp học máy cho dự đoán quãng đường và chỉ ra rằng việc sử dụng dữ liệu kỹ thuật có thể giảm phụ thuộc vào dữ liệu vận hành phức tạp. Gao và cộng sự [7] áp dụng Random Forest để ước lượng phạm vi hoạt động của EV dựa trên dung lượng pin, công suất sạc và trọng lượng xe, đạt được sai số dưới 10%. Một nghiên cứu khác của Zhang et al. [8] cho thấy XGBoost có khả năng khai thác tốt các quan hệ phi tuyến giữa đặc trưng và quãng đường.

Mặc dù đã có nhiều công bố, phần lớn tập trung vào dữ liệu vận hành (driving cycles) hoặc dữ liệu cảm biến. Số lượng công bố khai thác dữ liệu từ bảng thông số kỹ thuật công khai (ví dụ: dung lượng pin, kích thước xe, số chỗ, hệ dẫn động) vẫn còn hạn chế. Do đó, bài báo này hướng tới việc lấp đầy khoảng trống đó, bằng cách phát triển một khung học máy cho dự đoán quãng đường từ dữ liệu kỹ thuật.

III. PHƯƠNG PHÁP ĐỀ XUẤT

3.1. Tập dữ liệu

Bộ dữ liệu sử dụng trong nghiên cứu được tổng hợp từ cơ sở dữ liệu công khai về thông số kỹ thuật của các mẫu

xe điện thương mại hóa trên thị trường tính đến năm 2025 được công bố công khai trên nền tảng Kaggle [20]. Tập dữ liệu bao gồm 478 mẫu xe với 20 thuộc tính đầu vào được mô tả trong bảng 1. Dữ liệu được chia thành hai tập: 80% cho huấn luyện và 20% cho kiểm thử.

BẢNG 1. CÁC THUỘC TÍNH KỸ THUẬT ĐƯỢC DÙNG LÀM THUỘC TÍNH CHO MÔ HÌNH HỌC MÁY

STT	Tham số (thuộc tính)	Đơn vị/kiểu dữ liệu	Ghi chú
1	Dung lượng pin	kWh	Tổng dung tích lưu trữ năng lượng
2	Hiệu suất tiêu thụ năng lượng	Wh/km	Điện năng tiêu hao cho 1 km
3	Công suất cực đại	kW	Công suất tối đa động cơ
4	Mô-men xoắn cực đại	Nm	Lực xoắn tối đa
5	Tốc độ tối đa	Km/h	Tốc độ cực đại đạt được
6	Thời gian tăng tốc (0-100)	Giây	Thời gian từ 0 - 100km/h
7	Chiều dài	Mm	Kích thước hình học
8	Chiều rộng	Mm	Kích thước hình học
9	Chiều cao	Mm	Kích thước hình học
10	Trọng lượng	Kg	Tổng khối lượng xe
11	Số chỗ ngồi	Số nguyên	Sức chứa hành khách
12	Hệ dẫn động	Phân loại (FWD/RWD/AWD)	Kiểu chuyển động
13	Kiểu thân xe	Phân loại	Sedan, SUV, hatchback
14	Tầm hoạt động (range)	Km	Quãng đường di chuyển 1 lần sạc
15	Hiệu suất sạc (charging efficiency)	%	Hiệu suất chuyển đổi khi sạc
16	Công suất sạc nhanh tối đa	kW	Công suất sạc DC
17	Thời gian sạc chuẩn	Giờ	Thời gian với AC charger
18	Loại pin	Phân loại	LFP, NMC...
19	Hệ số cản gió (Cd)	Số thực	Ảnh hưởng đến khí động học
20	Kích thước mâm/lốp	Inch	Ảnh hưởng tiêu hao năng lượng

Biến mục tiêu là quãng đường di chuyển sau một lần sạc, được công bố theo các tiêu chuẩn kiểm định (WLTP hoặc EPA). Việc lựa chọn tập dữ liệu này cho phép khai thác các thông tin kỹ thuật sẵn có, phù hợp với mục tiêu xây dựng mô hình dự đoán quãng đường dựa trên thông số thiết kế của xe điện [1], [9].

3.2. Tiền xử lý dữ liệu

Trước khi huấn luyện mô hình, dữ liệu được tiền xử lý nhằm đảm bảo tính toàn vẹn và chất lượng đầu vào. Các giá trị khuyết thiếu được xử lý bằng cách loại bỏ nếu tỷ lệ thiếu vượt ngưỡng, hoặc thay thế bằng các phương pháp ước lượng thống kê (trung bình, trung vị) [10]. Để tránh sai lệch do khác biệt thang đo, các thuộc tính liên tục (ví dụ: dung lượng pin, công suất, hiệu suất năng lượng) được chuẩn hóa bằng kỹ thuật Min-Max scaling [11].

Các biến phân loại (chẳng hạn như hệ dẫn động, kiểu thân xe) được mã hóa bằng One-Hot Encoding nhằm chuyển đổi sang dạng số mà không làm mất thông tin về quan hệ giữa các lớp [12]. Ngoài ra, dữ liệu được kiểm tra và loại bỏ bản ghi trùng lặp, đồng thời phát hiện ngoại lai bằng phương pháp dựa trên khoảng tứ phân vị (IQR). Các bước này giúp giảm nhiễu và nâng cao hiệu quả của quá trình huấn luyện.

3.3. Thuật toán học máy

Trong nghiên cứu này, bốn thuật toán học máy được lựa chọn: Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), và K-Nearest Neighbors (KNN).

3.3.1. Decision Tree

DT là mô hình học có giám sát đơn giản nhưng trực quan, phân chia không gian đặc trưng dựa trên các ngưỡng tối ưu nhằm giảm thiểu độ bất thuần. Thuật toán DT được xây dựng dựa trên việc phân chia không gian đặc trưng một cách đệ quy thành các vùng rời rạc và gán một giá trị hoặc lớp đầu ra cho mỗi vùng. Quá trình này nhằm mục đích tìm ra một cấu trúc cây tối ưu, trong đó các nút bên trong đại diện cho các phép thử trên một đặc trưng, các nhánh đại diện cho kết quả của phép thử, và các nút lá đại diện cho kết quả dự đoán [13].

Xét một tập huấn luyện $D = \{(x_i, y_i)\}_{i=1}^N$, trong đó $x_i \in R^d$ là véc-tơ đặc trưng và là nhãn đầu ra. Cây quyết định tìm cách phân hoạch không gian đặc trưng R^d thành M vùng không chồng lấn R_1, R_2, \dots, R_M . Đối với một điểm x_i nằm trong vùng R_m , đầu ra dự đoán là \hat{y}_m .

Mô hình dự đoán $\hat{y}_m = \hat{f}(x)$ có dạng:

$$\hat{f}(x) = \sum_{m=1}^M c_m \cdot I(x \in R_m) \quad (1)$$

Quá trình xây dựng cây là tìm kiếm điểm phân tách tối ưu tại mỗi nút. Tiêu chí phân tách dựa trên việc giảm thiểu phương sai hoặc tổng bình phương Sai số (Sum of Squared Errors - SSE) trong các nút con. Giả sử tập được phân tách thành $D_{trái}$ và $D_{phải}$. Thuật toán tìm phân tách dựa trên các đầu vào và đầu ra (feature, threshold) sao cho tổng SSE của các tập con là nhỏ nhất, nghĩa là:

$$\min_{feature, threshold} \left(\sum_{x_i \in D_{trái}} (y_i - \bar{y}_{trái})^2 + \sum_{x_i \in D_{phái}} (y_i - \bar{y}_{phái})^2 \right) \quad (2)$$

trong đó $\bar{y}_{trái}$ và $\bar{y}_{phái}$ là giá trị trung bình của nhãn (đầu ra) trong các nút con tương ứng.

3.3.2. Random Forest

RF là phương pháp học tổ hợp (ensemble learning) dựa trên nhiều cây quyết định ngẫu nhiên, giúp giảm phương sai và cải thiện độ chính xác [14]. Ý tưởng cốt lõi của RF là huấn luyện nhiều cây quyết định độc lập trên các tập dữ liệu con được lấy mẫu ngẫu nhiên và sau đó tổng hợp kết quả dự đoán của chúng. Giả sử tập dữ liệu huấn luyện được cho bởi:

$$D = \{(x_i, y_i) | x_i \in R^d, y_i \in R, i=0, 1, \dots, N\} \quad (3)$$

trong đó $x_i = [x_{i1}, x_{i2}, \dots, x_{id}]^T$ là vector đặc trưng đầu vào, và y_i là giá trị đầu ra tương ứng.

Thuật toán RF bao gồm T cây quyết định độc lập, ký hiệu là $\{h_t(x), t=1,2,\dots,T\}$. Mỗi cây $h_t(x)$ được huấn luyện trên một tập con dữ liệu D_t được sinh ra bằng phương pháp lấy mẫu bootstrap từ tập dữ liệu gốc D . Tại mỗi nút của cây, RF chọn ngẫu nhiên một tập con gồm m đặc trưng ($m < d$) để tìm đặc trưng tối ưu cho việc chia nút, thay vì xét toàn bộ d đặc trưng. Điều này giúp giảm tương quan giữa các cây và tăng khả năng khái quát của mô hình. Sau khi huấn luyện, đầu ra của mô hình RF được xác định bằng cách tổng hợp kết quả của tất cả các cây trong rừng. Cụ thể:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x) \quad (4)$$

Tính ngẫu nhiên trong việc lấy mẫu dữ liệu và chọn đặc trưng giúp RF giảm thiểu hiện tượng overfitting so với các mô hình cây đơn lẻ. Ngoài ra, RF còn có khả năng đánh giá mức độ quan trọng của đặc trưng dựa trên mức giảm độ bất thuần (impurity decrease) hoặc dựa trên lỗi dự đoán khi hoán đổi đặc trưng [14].

3.3.3. Support Vector Machine

Thuật toán SVM có thể được mở rộng để giải quyết các bài toán hồi quy thông qua mô hình Support Vector Regression (SVR). Trong bài toán dự đoán quãng đường đi được của xe điện, SVR được sử dụng nhằm ước lượng quãng đường di chuyển tối đa \hat{R} dựa trên các đặc trưng đầu vào là các thông số kỹ thuật của xe điện. Giả sử tập dữ liệu huấn luyện được cho bởi:

$$D = \{(x_i, R_i) | x_i \in \mathbb{R}^d, R_i \in \mathbb{R}, i=0,1,\dots,N\} \quad (5)$$

trong đó $x_i = [x_{i1}, x_{i2}, \dots, x_{id}]^T$ là vector đặc trưng đầu vào biểu diễn các yếu tố ảnh hưởng (ví dụ: dung lượng pin, Hiệu suất tiêu thụ năng lượng, Công suất cực đại, Mô-men xoắn cực đại, v.v...), và R_i là quãng đường thực tế mà xe điện có thể di chuyển tương ứng với mẫu thứ i . Hàm hồi quy của mô hình SVR được biểu diễn dưới dạng:

$$f(x) = w^T \phi(x) + b \quad (6)$$

trong đó: $\phi(x)$ là hàm ánh xạ phi tuyến chuyển dữ liệu từ không gian đầu vào sang không gian đặc trưng, w là vector trọng số, và b là hệ số dịch (bias term).

Mục tiêu của SVR là tìm hàm $f(x)$ sao cho sai số tuyệt đối giữa giá trị dự đoán và giá trị thực R_i không vượt quá một ngưỡng cho phép ε , đồng thời đảm bảo độ phức tạp của mô hình là nhỏ nhất. Bài toán tối ưu của SVR được phát biểu như sau:

$$\min_{w,b,\xi_i,\xi_i^*} \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \right) \quad (7)$$

thỏa mãn:

$$R_i - (w^T \phi(x_i) + b) \leq \varepsilon + \xi_i$$

$$(w^T \phi(x_i) + b) - R_i \leq \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0, i=0,\dots,N$$

trong đó:

- ε là biên dung sai (epsilon-insensitive zone) xác định ngưỡng sai số cho phép,

- ξ_i, ξ_i^* là biến nới lỏng (slack variables) dùng để xử lý các điểm dữ liệu nằm ngoài biên ε ,

- $C > 0$ là tham số điều chuẩn (regularization parameter) điều chỉnh sự đánh đổi giữa độ chính xác huấn luyện và độ trơn của mô hình.

Sau khi giải bài toán tối ưu, hàm dự đoán của mô hình SVR có dạng:

$$\hat{R} = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \cdot K(x_i, x) + b \quad (8)$$

trong đó α_i và α_i^* là nhân tử Lagrange, $K(x_i, x) = \langle \phi(x_i), \phi(x) \rangle$ là hàm nhân (kernel function) mô tả độ tương đồng giữa các mẫu.

Một số hàm nhân thường được sử dụng trong bài toán dự đoán quãng đường của xe điện bao gồm:

(i) Tuyến tính: $K(x_i, x_j) = x_i^T x_j$

(ii) Đa thức: $K(x_i, x_j) = (x_i^T x_j + 1)^p$

(iii) Gaussian RBF: $K(x_i, x_j) = \exp\left[-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right]$

Việc lựa chọn tối ưu các siêu tham số của SVR (như C , ε , và σ) có vai trò quan trọng nhằm đạt được hiệu năng dự đoán cao và khả năng khái quát tốt cho các điều kiện vận hành khác nhau [15]. Trong nghiên cứu này, hàm nhân Gaussian RBF được lựa chọn do khả năng biểu diễn mối quan hệ phi tuyến giữa các yếu tố ảnh hưởng và quãng đường di chuyển của xe điện.

3.3.4. K-Nearest Neighbors

KNN là phương pháp dựa trên khoảng cách, trong đó giá trị dự đoán của mẫu mới được xác định từ các láng giềng gần nhất. KNN đơn giản và dễ triển khai, nhưng hiệu năng phụ thuộc mạnh vào tham số k và độ đo khoảng cách [16].

Quá trình dự đoán \hat{y} được thực hiện qua ba bước cơ bản:

Bước 1. Tính toán khoảng cách: Xác định K láng giềng gần nhất bằng cách sử dụng một hàm khoảng cách, phổ biến nhất là khoảng cách Euclidean giữa điểm truy vấn x_q và mọi điểm huấn luyện x_i :

$$d(x_q, x_i) = \sqrt{\sum_{j=1}^d (x_{qj} - x_{ij})^2} \quad (9)$$

Bước 2. Lựa chọn K láng giềng: Chọn tập hợp gồm K điểm huấn luyện có khoảng cách $d(x_q, x_i)$ nhỏ nhất.

Bước 3. Dự đoán giá trị hồi quy: Giá trị dự đoán là giá trị trung bình của các nhãn tương ứng của láng giềng. Công thức hồi quy cơ bản là:

$$\hat{y} = \frac{1}{K} \sum_{(x_i, y_i) \in N_K} y_i \quad (10)$$

Trong các ứng dụng đòi hỏi độ chính xác cao hơn, hồi quy có trọng số được ưu tiên sử dụng. Trong đó, mỗi láng

giếng được gán một trọng số (thường là nghịch đảo của khoảng cách) để ưu tiên các điểm gần hơn, khi đó giá trị ước lượng cho đầu ra được xác định bởi:

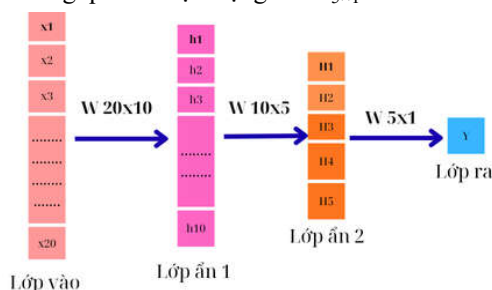
$$\hat{y} = \frac{\sum_{(x_i, y_i) \in N_K} \omega_i y_i}{\sum_{(x_i, y_i) \in N_K} \omega_i} \quad (11)$$

Sự kết hợp của bốn thuật toán này cho phép so sánh toàn diện giữa các cách tiếp cận khác nhau, từ mô hình dựa trên cây quyết định, học tổ hợp, đến phương pháp kernel và dựa trên khoảng cách.

3.4. Mô hình mạng nơ ron nhân tạo cho bài toán ước lượng quãng đường đi được của xe ô tô điện

Mạng nơ ron nhân tạo (Artificial Neural Network-ANN) là một trong những phương pháp học máy mạnh mẽ, được thiết kế mô phỏng cơ chế hoạt động của hệ thần kinh sinh học nhằm học và biểu diễn các quan hệ phi tuyến phức tạp giữa dữ liệu đầu vào và đầu ra. ANN có khả năng tự động trích xuất đặc trưng (feature extraction) và tổng quát hóa mô hình, nhờ đó được ứng dụng rộng rãi trong các bài toán dự đoán, nhận dạng và điều khiển phi tuyến [17]. Trong lĩnh vực phương tiện điện, ANN đặc biệt hữu ích cho việc ước lượng quãng đường di chuyển khả dụng (driving range estimation), vốn phụ thuộc vào nhiều yếu tố biến thiên như dung lượng pin, hành vi lái xe, tải trọng, điều kiện thời tiết và đặc tính đường. Không giống các mô hình vật lý truyền thống, ANN có khả năng học trực tiếp từ dữ liệu là thông số kỹ thuật được cung cấp từ nhà sản xuất xe điện giúp cải thiện đáng kể độ chính xác trong dự đoán phạm vi hoạt động của xe điện [18]. Nhờ khả năng mô hình hóa mạnh mẽ và linh hoạt, ANN trở thành nền tảng quan trọng cho việc phát triển các hệ thống quản lý năng lượng thông minh và tối ưu chiến lược vận hành trong xe ô tô điện.

Hình 1 minh họa kiến trúc của mạng ANN đề xuất cho bài toán ước lượng quãng đường di chuyển của xe ô tô điện. Mạng gồm ba tầng chính: lớp vào, hai lớp ẩn và lớp ra. Cụ thể, lớp vào bao gồm 20 đặc trưng đầu vào (x_1, x_2, \dots, x_{20}), có thể biểu diễn các thông số vận hành của xe được mô tả trong Bảng 1. Lớp ẩn thứ nhất gồm 10 nơ ron, được kết nối đầy đủ với lớp vào thông qua ma trận trọng số $W_{20 \times 10}$. Tiếp theo, lớp ẩn thứ hai gồm 5 nơ ron, nhận đầu vào từ lớp ẩn thứ nhất thông qua ma trận trọng số $W_{10 \times 5}$. Cuối cùng, lớp ra chỉ gồm một nơ ron duy nhất, đại diện cho giá trị ước lượng quãng đường di chuyển (Y), được tính toán từ lớp ẩn thứ hai thông qua ma trận trọng số $W_{5 \times 1}$.



Hình 1. Kiến trúc mạng ANN đề xuất

Các lớp ẩn có thể sử dụng hàm kích hoạt phi tuyến ReLU hoặc nhằm tăng khả năng biểu diễn các quan hệ phi tuyến phức tạp giữa các biến đầu vào. Mạng được huấn luyện bằng thuật toán lan truyền ngược để tối thiểu hóa hàm mất mát Mean Squared Error (MSE), nhằm cải thiện độ chính xác của mô hình trong quá trình dự đoán quãng đường thực tế. Kiến trúc hai lớp ẩn với số lượng nơ ron trung bình này được lựa chọn nhằm cân bằng giữa khả năng học biểu diễn và độ phức tạp tính toán của mô hình, phù hợp với bài toán hồi quy trong ước lượng quãng đường xe điện.

Với một mẫu đầu vào $x \in \mathbb{R}^{20}$, quá trình lan truyền tiến (forward propagation) được mô tả như sau:

$$h^{(1)} = \varphi^{(1)}(W^{(1)}x + b^{(1)}), \quad W^{(1)} \in \mathbb{R}^{10 \times 20},$$

$$h^{(2)} = \varphi^{(2)}(W^{(2)}h^{(1)} + b^{(2)}), \quad W^{(2)} \in \mathbb{R}^{5 \times 10},$$

$$\hat{y} = W^{(3)}h^{(2)} + b^{(3)}, \quad W^{(3)} \in \mathbb{R}^{1 \times 5}$$

trong đó $\varphi^{(1)}(\cdot)$ và $\varphi^{(2)}(\cdot)$ là các hàm kích hoạt phi tuyến, trong đề xuất này chúng tôi sử dụng hàm ReLU.

Hàm mất mát được sử dụng là MSE nhằm tối thiểu hoá sai số giữa giá trị dự đoán \hat{y}_i và giá trị thực y_i :

$$L = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (12)$$

Các trọng số $W^{(l)}$ và bias $b^{(l)}$ được tối ưu bằng thuật toán Adam để đạt được hội tụ nhanh và giảm thiểu overfitting trên tập huấn luyện. Mô hình sau khi huấn luyện được sử dụng để ước lượng quãng đường di chuyển còn lại của xe điện dựa trên các đặc trưng đầu vào, và hiệu năng của nó sẽ được đánh giá chi tiết trong phần mô phỏng tiếp theo.

3.5. Tiêu chí đánh giá mô hình

Các mô hình được đánh giá dựa trên ba chỉ số phổ biến trong các bài toán hồi quy: Sai số tuyệt đối trung bình (MAE), Căn bậc hai sai số bình phương trung bình (RMSE), và hệ số xác định (R^2). MAE phản ánh mức sai số trung bình trực tiếp, RMSE nhấn mạnh vào các sai số lớn, trong khi R^2 thể hiện mức độ giải thích của mô hình đối với biến mục tiêu [19].

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (13.a)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (13.b)$$

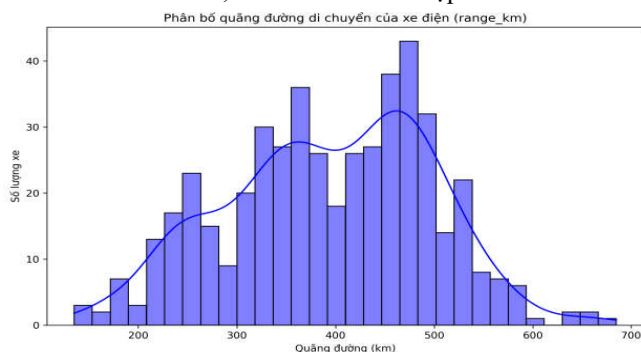
$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (13.c)$$

Trong đó: y_i là giá trị thực tế, \hat{y}_i là giá trị dự đoán, N là số mẫu, $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$ là giá trị trung bình của dữ liệu thực tế.

IV. KẾT QUẢ VÀ THẢO LUẬN

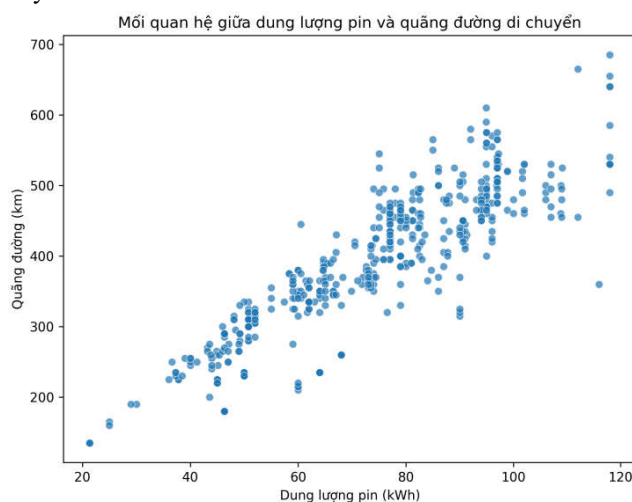
Mục này trình bày kết quả thực nghiệm sau khi áp dụng các phương pháp học máy và mô hình mạng nơ ron

nhân tạo để ước lượng quãng đường di chuyển của ô tô điện. Nội dung tập trung vào ba phần chính: (1) phân tích và trực quan hóa tập dữ liệu nhằm làm rõ phân bố của biến mục tiêu và mối tương quan giữa các đặc trưng đầu vào như dung lượng pin và quãng đường; (2) đánh giá quá trình huấn luyện mô hình ANN thông qua đường cong mất mát; và (3) so sánh hiệu năng dự đoán của năm mô hình (KNN, Random Forest, SVM, Decision Tree và ANN) dựa trên các chỉ số MAE, RMSE và R^2 trên tập kiểm thử.



Hình 2. Phân bố quãng đường di chuyển của xe điện

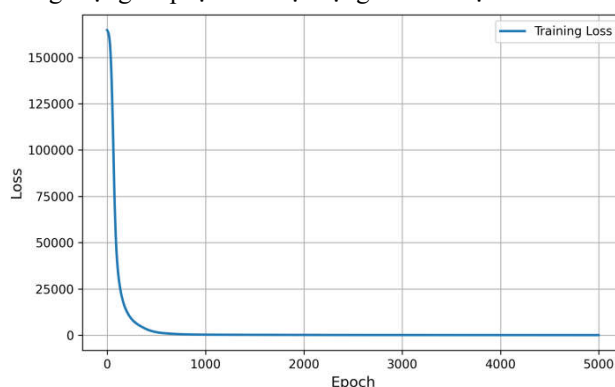
Biểu đồ histogram trong Hình 2 cho thấy phân bố quãng đường di chuyển sau một lần sạc của các mẫu xe điện có xu hướng lệch phải. Kết quả chỉ ra rằng phần lớn các mẫu xe tập trung trong khoảng 250 - 450 km, phản ánh mức phạm vi hoạt động phổ biến của các dòng xe điện trên thị trường hiện nay. Một số ít mẫu xe đạt quãng đường vượt quá 600 km, cho thấy sự hiện diện của các dòng xe cao cấp được trang bị pin dung lượng lớn. Đặc điểm phân bố có phần đuôi kéo dài về phía phải cho thấy sự không đồng đều giữa các phân khúc, từ các mẫu xe đô thị nhỏ gọn đến các mẫu xe cao cấp với khả năng di chuyển xa hơn.



Hình 3. Quan hệ giữa dung lượng pin và quãng đường di chuyển.

Biểu đồ phân tán trong Hình 3 thể hiện mối quan hệ giữa dung lượng pin và quãng đường di chuyển (range_km) của các mẫu xe điện. Kết quả cho thấy tồn tại mối quan hệ tuyến tính dương tương đối rõ rệt: dung lượng pin càng lớn thì quãng đường di chuyển càng xa. Tuy nhiên, các điểm dữ liệu cũng cho thấy sự phân tán

nhất định, phản ánh ảnh hưởng của các yếu tố khác như hiệu suất động cơ, thiết kế khí động học, trọng lượng xe và công nghệ quản lý năng lượng. Một số mẫu xe đạt được quãng đường di chuyển dài hơn so với mức trung bình cùng dung lượng pin, cho thấy sự khác biệt về công nghệ và tối ưu hóa hệ thống. Điều này khẳng định rằng ngoài dung lượng pin, các yếu tố kỹ thuật và thiết kế đóng vai trò quan trọng trong việc quyết định hiệu quả sử dụng năng lượng và phạm vi hoạt động của xe điện.



Hình 4. Đường biểu diễn giá trị mất mát của mô hình ANN

Hình 4 thể hiện đường hội tụ của hàm mất mát trong quá trình huấn luyện mô hình mạng nơ ron nhân tạo cho bài toán ước lượng quãng đường di chuyển của xe điện. Có thể quan sát thấy giá trị loss ban đầu rất lớn, sau đó giảm nhanh trong khoảng 500 epoch đầu tiên và dần ổn định khi số epoch tăng lên.

Điều này chứng tỏ quá trình huấn luyện đã hội tụ tốt: mô hình học được mối quan hệ giữa các đặc trưng đầu vào và đầu ra, đồng thời không xảy ra hiện tượng dao động hoặc tăng loss ở giai đoạn cuối-dấu hiệu cho thấy quá trình tối ưu đạt ổn định và không bị quá khớp (overfitting). Sau khoảng 2000 epoch, giá trị loss gần như không thay đổi đáng kể, cho thấy mô hình đã đạt trạng thái tối ưu.

Kết quả này khẳng định kiến trúc mạng và thuật toán tối ưu Adam được lựa chọn là phù hợp, đảm bảo khả năng hội tụ nhanh và độ chính xác cao trong dự đoán quãng đường còn lại của xe điện.

Mặc dù mối quan hệ giữa dung lượng pin và quãng đường di chuyển có xu hướng tuyến tính, các kết quả cho thấy ANN học được những quan hệ phi tuyến bậc cao giữa nhiều đặc trưng kết hợp. Cụ thể, cùng một dung lượng pin, quãng đường di chuyển có thể khác nhau đáng kể do ảnh hưởng đồng thời của hiệu suất tiêu thụ năng lượng (Wh/km), trọng lượng xe, hệ số cản gió và hệ dẫn động. ANN có khả năng mô hình hóa các tương tác phi tuyến này, ví dụ như mối quan hệ bù trừ giữa dung lượng pin và hiệu suất năng lượng, hoặc tác động phi tuyến của trọng lượng và kích thước xe lên mức tiêu thụ điện. Điều này giúp ANN đạt độ chính xác cao hơn so với các mô hình tuyến tính hoặc cây đơn lẻ, đặc biệt trong các vùng dữ liệu có sự phân tán lớn.

Mô hình mạng ANN và bốn mô hình học máy bao gồm DT, RF, SVM và KNN đã được triển khai để dự đoán quãng

đường di chuyển sau một lần sạc của xe điện. Kết quả đánh giá trên tập kiểm thử được trình bày trong Bảng 2.

BẢNG 2. HIỆU NĂNG DỰ ĐOÁN CỦA CÁC MÔ HÌNH HỌC MÁY

Mô hình	MAE (km)	RMSE (km)	R ²
KNN	22.70	30.98	0.889
Random Forest	11.17	15.04	0.974
SVM (SVR)	12.93	17.57	0.964
Decision Tree	15.57	20.39	0.952
ANN	12.9467	16.4192	0.9687

Bảng 2 trình bày kết quả so sánh hiệu năng của các mô hình học máy trong bài toán hồi quy ước lượng quãng đường di chuyển của xe điện, bao gồm KNN, Random Forest, SVR, Decision Tree và ANN.

Kết quả cho thấy Random Forest đạt hiệu năng tốt nhất với giá trị MAE = 11.17 km, RMSE = 15.04 km và hệ số xác định R²=0.974, thể hiện khả năng dự đoán chính xác và ổn định cao. Mô hình ANN cũng cho kết quả khả quan (MAE = 12.95 km, RMSE = 16.42 km, R²=0.9687), chỉ thấp hơn một chút so với Random Forest, chứng tỏ mạng nơ ron có thể học được mối quan hệ phi tuyến giữa các đặc trưng đầu vào và quãng đường di chuyển.

Trong khi đó, các mô hình SVM và Decision Tree đạt độ chính xác trung bình với R² lần lượt là 0.964 và 0.952. Mô hình KNN cho kết quả thấp nhất (MAE = 22.70 km, RMSE = 30.98 km, R²=0.889), cho thấy phương pháp dựa trên khoảng cách kém hiệu quả khi dữ liệu có tính phi tuyến và độ nhiễu cao.

Mặc dù mô hình ANN có khả năng học các quan hệ phi tuyến phức tạp, kết quả thực nghiệm cho thấy Random Forest (RF) đạt hiệu năng cao hơn một cách nhất quán trên tập kiểm thử. Nguyên nhân chính đến từ bản chất dữ liệu và đặc điểm của từng mô hình.

Thứ nhất, RF là mô hình ensemble dựa trên nhiều cây quyết định, có khả năng xử lý tốt dữ liệu tabular với số lượng mẫu vừa phải và nhiều đặc trưng không đồng nhất. Cơ chế lấy mẫu bootstrap và lựa chọn ngẫu nhiên tập đặc trưng tại mỗi nút giúp RF giảm phương sai và hạn chế overfitting, đặc biệt hiệu quả khi dữ liệu chứa nhiễu và mối quan hệ phi tuyến cục bộ.

Thứ hai, mô hình ANN trong nghiên cứu này có kiến trúc tương đối đơn giản và được huấn luyện trên tập dữ liệu có kích thước hạn chế (478 mẫu). Trong bối cảnh này, ANN khó khai thác hết tiềm năng biểu diễn so với RF, vốn không yêu cầu nhiều dữ liệu để đạt hiệu năng ổn định. Do đó, RF thể hiện ưu thế rõ rệt về độ chính xác và khả năng tổng quát hóa trong bài toán dự đoán quãng đường từ thông số kỹ thuật.

Dựa trên đặc điểm dữ liệu và các nghiên cứu trước đó, dung lượng pin và hiệu suất tiêu thụ năng lượng (Wh/km) là hai đặc trưng có ảnh hưởng mạnh nhất đến quãng đường di chuyển của xe điện. Ngoài ra, các yếu tố như trọng lượng xe, hệ số cản gió và kiểu thân xe cũng đóng vai trò quan trọng thông qua tác động gián tiếp lên mức tiêu hao năng lượng. Trong các mô hình ensemble như Random Forest, tầm quan trọng của đặc trưng thường được xác định thông qua mức giảm độ bất thuần khi phân tách nút.

Kết quả thực nghiệm cho thấy các đặc trưng liên quan đến pin và hiệu suất năng lượng đóng góp lớn nhất vào khả năng dự đoán, trong khi các đặc trưng hình học và cấu hình dẫn động đóng vai trò hỗ trợ.

Nhìn chung, kết quả thực nghiệm chứng minh rằng các mô hình học máy (như RF) và mô hình ANN có khả năng mô tả tốt mối quan hệ phức tạp giữa các yếu tố vận hành của xe điện và quãng đường di chuyển thực tế, từ đó có tiềm năng ứng dụng cao trong các hệ thống quản lý năng lượng thông minh.

Mặc dù kiến trúc ANN trong nghiên cứu này được thiết kế ở mức cơ bản với hai lớp ẩn, mục tiêu chính của bài báo không nhằm đề xuất một kiến trúc mạng nơ ron mới, mà tập trung vào việc đánh giá khả năng áp dụng ANN trong bối cảnh dữ liệu thông số kỹ thuật công khai, nơi kích thước dữ liệu còn hạn chế.

Việc sử dụng kiến trúc ANN đơn giản giúp giảm nguy cơ quá khớp và đảm bảo tính công bằng khi so sánh với các mô hình học máy truyền thống. Kết quả thực nghiệm cho thấy, ngay cả với cấu trúc cơ bản, ANN vẫn đạt hiệu năng cạnh tranh so với các mô hình tiên tiến như Random Forest và SVR. Điều này cho thấy tiềm năng mở rộng của ANN khi được kết hợp với các kỹ thuật nâng cao hơn như tối ưu siêu tham số, regularization, hoặc kiến trúc học sâu trong các nghiên cứu tiếp theo.

V. KẾT LUẬN

Trong nghiên cứu này, chúng tôi đã xây dựng và đánh giá các mô hình học máy để dự đoán quãng đường di chuyển của xe điện dựa trên các thông số kỹ thuật như dung lượng pin, công suất động cơ và hiệu suất năng lượng. Kết quả cho thấy các mô hình mạng nơ ron và học máy truyền thống có khả năng dự đoán tương đối chính xác, trong đó các mô hình ensemble như Random Forest đạt hiệu suất cao hơn. Phân tích dữ liệu xác nhận dung lượng pin là yếu tố ảnh hưởng mạnh nhất, song các đặc điểm thiết kế và quản lý năng lượng cũng đóng vai trò quan trọng.

Nghiên cứu vẫn còn một số hạn chế, bao gồm dữ liệu chủ yếu dựa trên thông số công bố, số lượng biến đầu vào hạn chế và chưa khai thác các mô hình học sâu hoặc lai ghép. Trong tương lai, việc tích hợp dữ liệu thực tế từ cảm biến, kết hợp các mô hình dựa trên vật lý, mở rộng tập dữ liệu và thử nghiệm các phương pháp học sâu hoặc meta-learning sẽ giúp cải thiện độ chính xác và khả năng tổng quát hóa. Những hướng đi này hứa hẹn nâng cao hiệu quả dự đoán, đồng thời hỗ trợ thiết kế và tối ưu hóa xe điện trong thực tiễn.

TÀI LIỆU THAM KHẢO

- [1] International Energy Agency, *Global EV Outlook 2023*.
- [2] Ehsani, M., Gao, Y., & Emadi, A., *Modern Electric, Hybrid Electric, and Fuel Cell Vehicles*. CRC Press, 2017. <https://doi.org/10.1201/9781420054002>.
- [3] K. Hu, J. Wu, and M. Liu, "Modelling of electric vehicles energy consumption from the perspective of field test data and driving style questionnaires," *China Simulation*, vol. 30, no. 11, pp. 1-7, 2018.

- [4] Ozkan, Mehmet Fatih, et al. "Data-Driven Personalized Energy Consumption Range Estimation for Plug-in Hybrid Electric Vehicles in Urban Traffic." *IFAC-PapersOnLine* 58.28 (2024): 162-167.
- [5] Xinfang Zhang, Zhe Zhang, Yang Liu, Zhigang Xu, Xiaobo Qu, A review of machine learning approaches for electric vehicle energy consumption modelling in urban transportation, *Renewable Energy*, Volume 234,2024.
- [6] Han, S., et al., "Deep learning based energy consumption prediction for electric vehicles," *Energy*, vol. 208, 2020.
- [7] Gao, Z., et al., "Electric vehicle range prediction using machine learning," *IEEE Access*, vol. 8, pp. 145051-145062, 2020.
- [8] Zhang, Y., et al., "Application of gradient boosting decision trees for electric vehicle range estimation," *Energies*, vol. 14, no. 5, 2021
- [9] Wang, Y., et al., "Machine learning approaches for electric vehicle range prediction: A review," *Renewable and Sustainable Energy Reviews*, vol. 148, 2021.
- [10] Little, R. J. A., & Rubin, D. B., *Statistical Analysis with Missing Data*, Wiley, 2019.
- [11] Jain, A., et al., "Data preprocessing for machine learning models," *Journal of Big Data*, vol. 7, no. 1, 2020.
- [12] Pedregosa, F., et al., "Scikit-learn: Machine learning in Python," *JMLR*, vol. 12, pp. 2825-2830, 2011.
- [13] Quinlan, J. R., *C4.5: Programs for Machine Learning*, Morgan Kaufmann, 1993.
- [14] Breiman, L., "Random forests," *Machine Learning*, vol. 45, pp. 5-32, 2001.
- [15] Smola, A. J., & Schölkopf, B., "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, pp. 199-222, 2004.
- [16] Cover, T., & Hart, P., "Nearest neighbor pattern classification," *IEEE Trans. Information Theory*, vol. 13, no. 1, pp. 21-27, 1967.
- [17] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed., Pearson, 2009.
- [18] J. V. Ferreira et al., "Electric vehicle driving range prediction using neural networks," *Energy Reports*, vol. 8, pp. 1253-1262, 2022.
- [19] Chai, T., & Draxler, R. R., "Root mean square error (RMSE) or mean absolute error (MAE)?-Arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, 2014.
- [20] <https://www.kaggle.com/datasets/urvishahir/electric-vehicle-specifications-dataset-2025>.