

ỨNG DỤNG THUẬT TOÁN KNN TRONG CHẨN ĐOÁN VÀ DỰ BÁO HƯ HỎNG HỘP SỐ TỰ ĐỘNG

APPLICATION OF KNN ALGORITHM FOR DIAGNOSIS AND PREDICTION OF AUTOMATIC TRANSMISSION FAILURES

PHẠM ANH TUẤN, VŨ NGỌC TUẤN*, NGUYỄN VĂN TRÀ, NGUYỄN VĂN TỬ

Viện Cơ khí động lực, Học viện Kỹ thuật quân sự

**Email liên hệ: tuan.vungoc@lqdtu.edu.vn*

DOI: <https://doi.org/10.65154/jmst.2025.i84.857>

Tóm tắt

Nghiên cứu này mô phỏng các dạng hư hỏng của hộp số tự động trong điều kiện vận hành của xe ô tô tại Việt Nam bằng phần mềm Simulation-X, qua đó xây dựng bộ dữ liệu về đặc tính vận hành và lỗi thường gặp. Trên cơ sở đó, hai nhóm mô hình học máy (Machine Learning) dựa trên thuật toán K-nearest neighbors (KNN) được phát triển: (1) mô hình phân loại hư hỏng với độ chính xác là 92,2% và (2) các mô hình hồi quy dự báo mức hao tổn cơ khí hoặc thủy lực với hệ số xác định R^2 xấp xỉ 1.0, vượt xa các mô hình hồi quy tuyến tính truyền thống. Kết quả đánh giá cho thấy các mô hình KNN có khả năng chẩn đoán và dự báo hiệu quả, đồng thời gợi ý tiềm năng ứng dụng trong giám sát thời gian thực, nâng cao khả năng chẩn đoán sớm và tối ưu hóa bảo dưỡng, sửa chữa hộp số tự động trên xe ô tô.

Từ khóa: Hộp số tự động, chẩn đoán, dự báo, KNN, phân loại, hồi quy.

Abstract

In this study, various failure types of automatic transmissions under real-world driving conditions in Vietnam were simulated using Simulation-X software, thereby constructing a dataset of operating characteristics and common failures. Based on this dataset, two groups of machine learning models using the K-nearest neighbors (KNN) algorithm were developed: (1) a classification model for failures diagnosis with the accuracy of 92,2% and (2) a regression model for predicting mechanical or hydraulic loss with determination coefficients (R^2) close to 1.0, significantly outperforming traditional linear regression models. Evaluation results demonstrate that the KNN models provide effective diagnosis and prediction, highlighting their potential application in real-time

monitoring, early failures detection, and optimization of maintenance and repair for automotive automatic transmissions.

Keywords: Automatic transmission, diagnose, prediction, KNN, classification, regression.

1. Tổng quan vấn đề nghiên cứu

Các hư hỏng của hộp số ô tô dẫn đến giảm hiệu suất, tăng chi phí bảo trì và gây mất an toàn trong quá trình vận hành. Hiện tượng mài mòn hoặc gãy bánh răng thường xảy ra khi mô-men xoắn vượt giới hạn thiết kế hoặc do thiếu bôi trơn, bảo trì (Gama và các cộng sự, 2023) [1]. Hiện nay, có nhiều phương pháp xác định hỏng hóc của hộp số tự động, từ kiểm tra cơ khí đến ứng dụng trí tuệ nhân tạo (AI), với các ưu nhược điểm riêng.

Hư hỏng của hộp số tự động thường liên quan đến ba hệ thống: Điện tử (phát hiện lỗi qua OBD), cơ khí (phanh, ly hợp - khó phát hiện qua OBD) và thủy lực (liên quan đến áp suất dầu). Nghiên cứu này tập trung vào mô hình chẩn đoán hỏng hóc cơ khí và mất mát thủy lực của hộp số tự động thông qua các tham số động lực học có thể đo lường được.

Các phương pháp truyền thống kiểm tra thủ công các bộ phận như ly hợp, bánh răng, phanh; tuy hiệu quả trong phát hiện lỗi rõ rệt nhưng tốn thời gian, chi phí cao và phụ thuộc kỹ năng của kỹ thuật viên. Gama và các cộng sự trong nghiên cứu của mình đã sử dụng thử nghiệm độ cứng Rockwell, kiểm tra thăm thâu và quan sát để phân tích hỏng hóc bánh răng. Chen và các cộng sự (2012) khảo sát phát hiện lỗi bằng phân tích hàm lượng kim loại trong dầu hộp số, giúp chẩn đoán hư hỏng theo phương pháp không xâm lấn nhưng với chi phí và yêu cầu chuyên môn cao [2].

Các hệ thống hiện đại như hệ thống chẩn đoán trên xe (OBD) phân tích mô-men xoắn, tốc độ, áp suất dầu để phát hiện lỗi. Merkisz (2008) cũng như Avinash và các cộng sự (2025) nghiên cứu các hệ thống OBD tiên tiến tích hợp nhiều giao thức truyền dữ liệu nhằm nâng cao khả năng chẩn đoán, quản lý đội xe và tối ưu bảo trì [3], [4].

Zhang và các cộng sự (2024) trong nghiên cứu của mình đã chẩn đoán hư hỏng dựa trên dao động sử dụng gia tốc kế và microphone ghi nhận tiếng ồn, dao động bất thường, áp dụng biến đổi GHM đa sóng và giải thuật giải liên hợp để phát hiện lỗi bánh răng với độ chính xác cao [5]. Zuñiga và các cộng sự (2024) sử dụng biến đổi Fourier nhanh để phân tích dao động bánh răng, trong khi Chengyun và các cộng sự (2024) kiểm tra chất lượng âm thanh hộp số trên băng thử [6], [7]. Các phương pháp này hiệu quả nhưng yêu cầu xử lý tín hiệu phức tạp.

Công nghệ trí tuệ nhân tạo (AI) trong những năm gần đây được áp dụng mạnh mẽ trong việc chẩn đoán hư hỏng trên xe ô tô. Aditya và các cộng sự (2024) sử dụng mạng nơ-ron nhân tạo để chẩn đoán lỗi ly hợp ma sát khô với độ chính xác cao [8]. Haibing và các cộng sự (2024) cải tiến mô hình mạng Nơ-ron tích chập (CNN) YOLOv8 nhằm phát hiện khuyết tật bề mặt bánh răng, nâng cao độ chính xác và tốc độ phát hiện [9]. Badiger (2024) phát triển mô hình dự đoán tuổi thọ dầu hộp số theo thời gian thực dựa trên đặc tính dầu và tham số vận hành, nhằm ngăn ngừa chi phí sửa chữa lớn và cải thiện hiệu suất [10].

So sánh các phương pháp cho thấy, trong khi kỹ

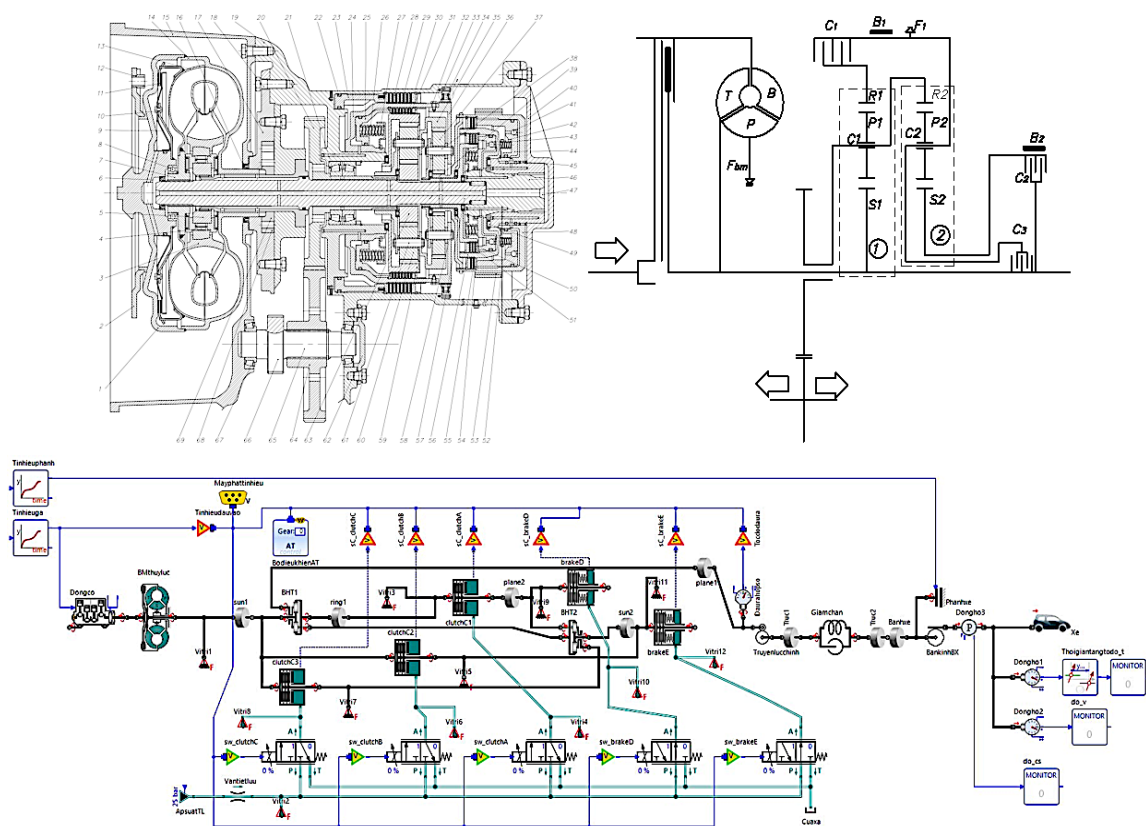
thuật truyền thống vẫn hữu ích và vẫn đang được áp dụng rộng rãi vì những ưu điểm của mình, các mô hình AI mang lại độ chính xác, khả năng tự động hóa và dự đoán vượt trội, hứa hẹn sẽ dần thay thế các phương pháp truyền thống.

Tuy nhiên, việc xây dựng mô hình chẩn đoán và dự báo hư hỏng cho hộp số tự động trong điều kiện vận hành đặc thù tại Việt Nam vẫn chưa được quan tâm đúng mức. Nghiên cứu này đóng góp bằng việc (i) thiết lập bộ dữ liệu mô phỏng toàn diện các dạng hư hỏng hộp số tự động bằng phần mềm Simulation-X, (ii) phát triển hai nhóm mô hình học máy dựa trên thuật toán KNN cho chẩn đoán và dự báo hư hỏng, và (iii) đánh giá khả năng ứng dụng trong giám sát thời gian thực và tối ưu hóa bảo dưỡng. Những điểm mới này giúp bổ sung khoảng trống nghiên cứu, đồng thời gợi mở hướng triển khai AI vào thực tiễn khai thác, bảo dưỡng xe ô tô với hộp số tự động tại Việt Nam.

2. Phương pháp nghiên cứu

2.1. Mô phỏng hư hỏng của hộp số tự động và tạo bộ dữ liệu cho nghiên cứu

Tại Việt Nam, các phương tiện sử dụng hộp số tự động chịu ảnh hưởng bởi nhiều yếu tố như khí hậu



Hình 1. Từ trên xuống dưới, từ trái sang phải: Mặt cắt dọc, sơ đồ động học và mô hình mô phỏng được xây dựng bằng phần mềm Simulation-X của hộp số tự động (C1, C2, C3 - ly hợp; B1, B2 - phanh)

nhật đới gió mùa với nhiệt độ cao, độ ẩm lớn, tạo điều kiện cho bụi bẩn xâm nhập gây mài mòn chi tiết cơ khí; cơ sở hạ tầng giao thông kém và điều kiện vận hành phức tạp, yêu cầu thay đổi tỷ số truyền liên tục; cùng với chế độ bảo dưỡng hạn chế. Các hư hỏng hộp số chủ yếu xuất phát từ hiện tượng mài mòn bánh răng, trục truyền, ly hợp; rò rỉ dầu hoặc sử dụng dầu kém chất lượng; và lỗi cảm biến, mô-đun điều khiển, phần mềm xử lý tín hiệu.

Trong nghiên cứu này, đối tượng khảo sát là hộp số tự động lắp trên xe KIA Morning Si AT đời 2016. Mô hình hộp số được xây dựng trên phần mềm Simulation-X, tái hiện đầy đủ cấu trúc truyền động, mô hình động học chuyển động tịnh tiến của xe và tương tác giữa lốp với mặt đường (Hình 1). Mô hình mô phỏng bao gồm ECU điều khiển chuyển số, các bộ truyền bánh răng hành tinh, hệ thống thủy lực, hệ thống phanh và ly hợp để điều khiển đóng/mở các cụm bánh răng, cũng như hệ thống thu thập dữ liệu đầu ra như thời gian tăng tốc, vận tốc cuối cùng và công suất đầu ra hộp số. Ngoài ra, các yếu tố lỗi như tổn thất cơ học tại các cụm phanh và ly hợp, rò rỉ dầu trong hệ thống thủy lực cũng được đưa vào mô phỏng nhằm phân tích, chẩn đoán và xác định nguyên nhân hư hỏng của hộp số tự động.

Kết quả mô phỏng cho thấy dạng hư hỏng đặc trưng của hộp số tự động được đặc trưng bởi ba thông số động học của xe như sau:

- *t100 (s): Thời gian tăng tốc của xe từ 0 đến 100 km/h;*

- *v30 (m/s): Vận tốc xe sau 30 giây;*

- *P30 (kW): Công suất của xe sau 30 giây.*

Biến đầu ra thể hiện dạng hư hỏng của hộp số tự động, được phân loại thành một trong 12 nhóm sau:

- *Normal: Điều kiện hoạt động bình thường, không có hao tổn cơ khí đáng kể và/hoặc không mất mát thủy lực;*

- *TorqueConverter: Hư hỏng bộ biến mô;*

- *MechanicClutchC1: Hao tổn cơ khí ly hợp C1;*

- *HydraulicClutchC1: Mất mát thủy lực ly hợp C1;*

- *MechanicClutchC2: Hao tổn cơ khí ly hợp C2;*

- *HydraulicClutchC2: Mất mát thủy lực ly hợp C2;*

- *MechanicClutchC3: Hao tổn cơ khí ly hợp C3;*

- *HydraulicClutchC3: Mất mát thủy lực ly hợp C3;*

- *MechanicBrakeB1: Hao tổn cơ khí phanh B1;*

- *HydraulicBrakeB1: Mất mát thủy lực phanh B1;*

- *MechanicBrakeB2: Hao tổn cơ khí phanh B2;*

- *HydraulicBrakeB2: Mất mát thủy lực phanh B2.*

Gán hao tổn cơ khí và mất mát thủy lực tại các ly hợp và phanh chuyển số vào 11 vị trí trên mô hình mô phỏng với các giá trị ngưỡng. Bằng cách tăng dần mức hao tổn ta thu được dữ liệu cuối cùng bao gồm năm cột với các thuộc tính sau:

- *failure: Biến phân loại đại diện cho loại hư hỏng, được phân vào một trong 12 danh mục;*

- *t100: Biến số liên tục đại diện cho thời gian tăng tốc từ 0 đến 100 km/h;*

- *v30: Biến số liên tục đại diện cho vận tốc xe sau 30 giây;*

- *P30: Biến số liên tục đại diện cho công suất xe sau 30 giây;*

- *L: Hao tổn cơ khí hoặc thủy lực tương ứng với loại hư hỏng.*

Bộ dữ liệu sau đó được xáo trộn ngẫu nhiên và chia thành 70% dùng cho huấn luyện và 30% dùng cho kiểm tra. Trong quá trình huấn luyện, một phần dữ liệu được dùng cho xác thực chéo 3-folds nhằm đánh giá khách quan độ chính xác. Do dữ liệu gốc có khoảng 1.000 điểm cho mỗi loại hư hỏng (tổng hơn 12.000 điểm), quy mô này vừa đảm bảo ý nghĩa thống kê vừa nâng cao độ tin cậy và độ chính xác của các mô hình.

2.2. Thuật toán KNN trong giải quyết bài toán phân loại và bài toán hồi quy

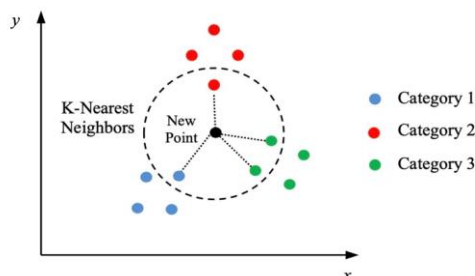
Trong những năm gần đây, học máy (Machine Learning) - một nhánh của trí tuệ nhân tạo - phát triển mạnh với nhiều thuật toán giải quyết hai bài toán cốt lõi: Phân loại và hồi quy. Phân loại gán mẫu dữ liệu vào các nhãn xác định trước (ở đây gồm 12 loại hư hỏng hộp số tự động, dựa trên ba biến đầu vào là t100, v30, P30), trong khi hồi quy phân tích mối quan hệ giữa biến phụ thuộc và độc lập để cho ra giá trị liên tục.

Trong nghiên cứu này, dựa trên ba tham số đầu vào là t100, v30, P30, thuật toán phân loại được sử dụng để xác định các hư hỏng như lỗi bộ biến mô, hư hỏng cơ khí hoặc thủy lực của ly hợp và phanh, còn thuật toán hồi quy được áp dụng để ước tính mức hao tổn cơ khí hoặc mất mát thủy lực (theo %).

Sau khi so sánh một số thuật toán học máy, nghiên cứu lựa chọn K-Nearest Neighbors (KNN) để phát triển mô hình phân loại và mô hình hồi quy cho hộp số tự động.

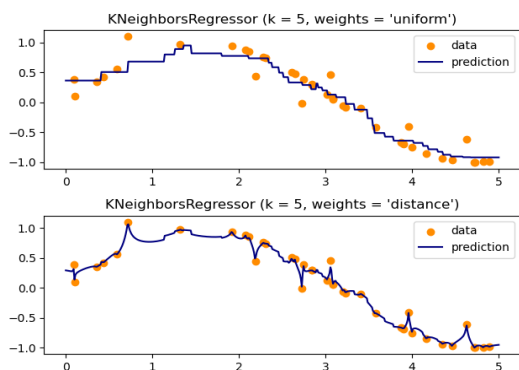
Mô hình KNN được minh họa đơn giản trong Hình 2 [11]. Giả sử có ba lớp trong không gian đặc trưng hai chiều, mỗi lớp được biểu diễn bằng một màu khác nhau. Với một điểm chưa xác định, thuật toán xét các "láng giềng gần nhất"; điểm này sẽ được phân loại vào lớp có số láng giềng chiếm đa số, chẳng hạn thuộc lớp

màu xanh lá cây (loại 3) khi có 2 trong 4 láng giềng gần nhất. Giá trị k quyết định số láng giềng được xem xét (ví dụ $k = 1$ thì điểm sẽ nhận lớp của láng giềng gần nhất). Độ chính xác phân loại được tính bằng tỷ lệ giữa số dự đoán đúng và tổng số dự đoán.



Hình 2. Mô tả nguyên lý thuật toán phân loại KNN

Ngoài việc sử dụng để giải bài toán phân loại, thuật toán KNN có thể được sử dụng để giải bài toán hồi quy trong các trường hợp nhãn dữ liệu là các biến liên tục thay vì biến rời rạc. Nhận được gán cho điểm truy vấn được tính dựa trên giá trị trung bình của các nhãn từ các láng giềng gần nhất, như minh họa trong Hình 3 [12].



Hình 3. Mô tả nguyên lý thuật toán hồi quy KNN

Các tác giả sử dụng thư viện mã nguồn mở Scikit-Learn để phân tích dữ liệu, huấn luyện và tối ưu hóa mô hình. Với thuật toán KNN, số láng giềng k quyết định độ phức tạp của mô hình: k lớn giúp giảm phương sai nhưng dễ gây “Underfitting”, trong khi k nhỏ tạo ranh giới quyết định cục bộ hơn nhưng có nguy cơ “Overfitting” [12]. Các giá trị k được khảo sát trong tập:

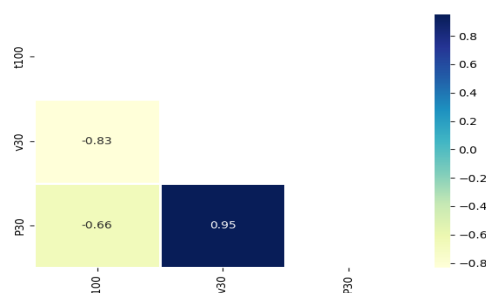
$k \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 15, 17, 20, 25, 30, 40, 60, 80, 100\}$.

Quá trình xây dựng mô hình có điều chỉnh siêu tham số này nhằm tối ưu hóa độ chính xác phân loại lỗi hộp số tự động, đồng thời nâng cao hiệu quả dự báo hao tổn cơ khí và mất mát thủy lực.

3. Kết quả nghiên cứu

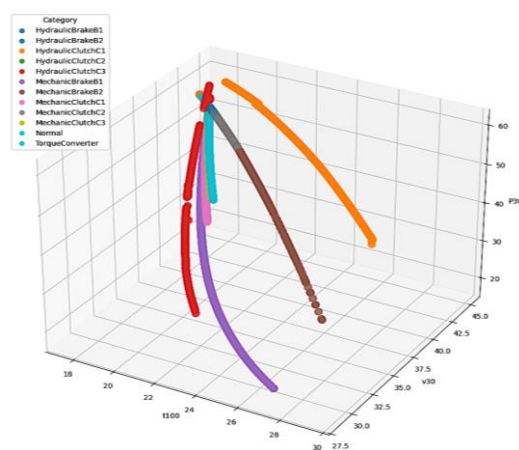
3.1. Giai đoạn tiền xử lý dữ liệu

Trong giai đoạn tiền xử lý, các điểm dữ liệu phát sinh lỗi trong mô phỏng đã bị loại bỏ nhằm bảo đảm độ tin cậy của mô hình. Tập dữ liệu sau lọc được kiểm tra phân bố biến và tiến hành phân tích khám phá dữ liệu (EDA) bằng hệ số tương quan Pearson để đánh giá mối quan hệ giữa các đặc trưng. Mục tiêu là nhận diện các cặp biến có tương quan mạnh, có nguy cơ gây đa cộng tuyến và làm sai lệch kết quả mô hình thống kê, nhất là trong các phương pháp hồi quy. Kết quả phân tích được trực quan hóa qua bản đồ nhiệt tương quan (Hình 4) để làm cơ sở lựa chọn và loại bỏ đặc trưng dư thừa.



Hình 4. Ma trận tương quan các biến đặc trưng

Trong nghiên cứu này, các tác giả giữ lại toàn bộ các đặc trưng cho phân tích tiếp theo mặc dù quan sát thấy mối tương quan cao giữa $v30$ và $P30$, nhằm giữ tối đa thông tin. Quyết định cuối cùng về việc loại bỏ đặc trưng sẽ được đưa ra sau khi có kết quả phân tích độ quan trọng hoán vị (Permutation Importance).



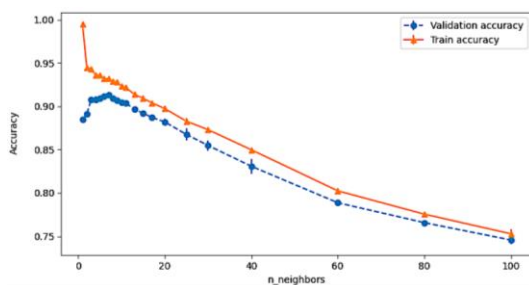
Hình 5. Minh họa không gian đặc trưng phân loại

Trong không gian ba chiều với ba biến đầu vào ($t100$, $v30$, $P30$), việc trực quan hóa (Hình 5) cho thấy 12 lớp hư hỏng hộp số tự động phần lớn được tách biệt nhưng vẫn tồn tại một số vùng chồng lấn, đặc biệt giữa

các trạng thái mất công suất và rò rỉ thủy lực nhẹ, gợi ý nguy cơ phân loại nhằm do ranh giới quyết định không rõ ràng. Phần tiếp theo sẽ trình bày kết quả huấn luyện và điều chỉnh siêu tham số cho mô hình KNN trong phân loại các loại hư hỏng, cũng như xây dựng mô hình KNN hồi quy để dự báo tổn thất cơ khí hoặc mất mát thủy lực.

3.2. Xây dựng mô hình phân loại để chẩn đoán hư hỏng của hộp số tự động

Trong quá trình tối ưu siêu tham số, độ chính xác xác thực (Validation Accuracy) của mô hình KNN đạt cực đại ở 91,3% khi số lượng láng giềng tối ưu (k) bằng 7, như minh họa trong Hình 6.



Hình 6. Kết quả tối ưu siêu tham số k khi phân loại

Khi số láng giềng k tăng lên vượt quá 7, độ chính xác của mô hình giảm do hiện tượng “Underfitting”. Các tác giả sử dụng độ chính xác xác thực (Validation Accuracy) làm chỉ số chính trong quá trình tinh chỉnh mô hình để đảm bảo đánh giá đáng tin cậy về khả năng khái quát hóa của KNN trên dữ liệu mới.

Sau khi xác định được mô hình KNN tối ưu với số hàng xóm tốt nhất ($k = 7$), mô hình được đánh giá trên tập dữ liệu kiểm tra. Trong trường hợp này, độ chính xác trên tập kiểm tra của mô hình KNN đạt 92,2%.

Phân tích ma trận nhầm lẫn chuẩn hóa (Hình 7) đã bổ sung thêm thông tin chi tiết về độ chính xác phân loại của mô hình trên tất cả các loại hỏng hóc hộp số ô tô. Phân tích này cho thấy mô hình KNN đạt độ chính xác phân loại cao, khiến nó thành phương pháp đầy tiềm năng để phân loại các hỏng hóc hộp số ô tô dựa trên ba đặc trưng đầu vào ($t100$, $v30$ và $P30$). Trong quá trình kiểm tra, mô hình KNN cho kết quả phân loại hoàn hảo (độ chính xác 100%) đối với ba nhóm sau:

- Mất mát thủy lực ở ly hợp C1;
- Hao tổn cơ khí ở phanh B1;
- Hao tổn cơ khí ở ly hợp C1.

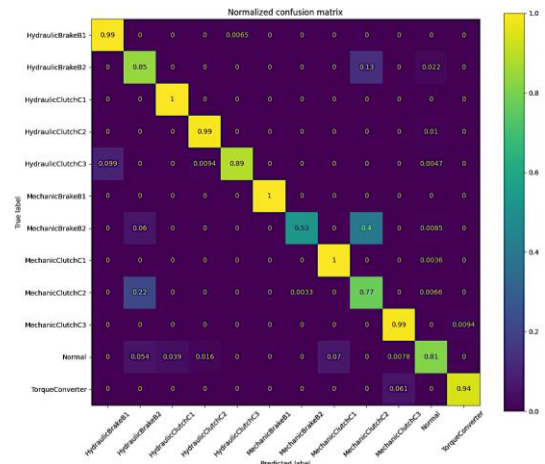
Ngoài ra, với một số nhầm lẫn nhỏ, mô hình KNN vẫn có thể phân loại đáng tin cậy (độ chính xác 99%):

- Mất mát thủy lực ở phanh B1;

- Mất mát thủy lực ở ly hợp C2;

- Hao tổn cơ khí ở ly hợp C3.

Mặc dù vậy, mô hình gặp khó khăn với hao tổn cơ khí ở phanh B2 khi chỉ đạt độ chính xác 53%, chủ yếu bị nhầm thành hao tổn cơ khí của ly hợp C2 (40%).



Hình 7. Ma trận nhầm lẫn chuẩn hóa của mô hình

3.3. Xây dựng các mô hình hồi quy dự báo hao tổn cơ khí hoặc mất mát thủy lực của hư hỏng

Trên cơ sở kết quả phân loại hư hỏng và xác định vị trí xảy ra hư hỏng, các tác giả tiến hành xây dựng các mô hình hồi quy sử dụng thuật toán KNN nhằm dự báo hao tổn cơ khí (hoặc mất mát thủy lực) L tại vị trí đã xác định.

Việc xây dựng các mô hình hồi quy có thể được thực hiện bằng thư viện Scikit-Learn như đối với trường hợp xây dựng các mô hình phân loại ở trên. Ngoài ra có nhiều phần mềm tích hợp sẵn chức năng xây dựng các mô hình này, trong bài viết các tác giả sử dụng phần mềm mã nguồn mở JASP để xây dựng và tối ưu mô hình, đồng thời xây dựng các mô hình tuyến tính để làm cơ sở so sánh.

Với mỗi loại hư hỏng, các tác giả tiến hành xây dựng mô hình và tối ưu siêu tham số k , với mục tiêu là giảm thiểu tối đa sai số trung bình bình phương (Mean Square Error - MSE).

Kết quả xây dựng mô hình hồi quy xác định giá trị mất mát thủy lực tại vị trí phanh B1 (Hình 8), cho thấy:

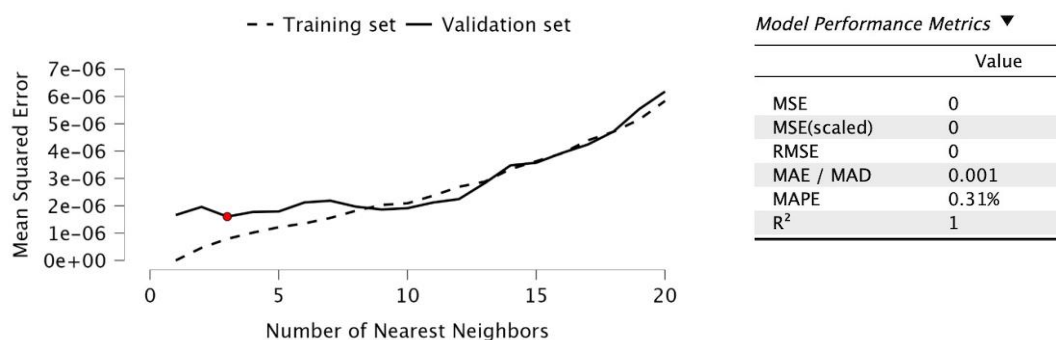
- Giá trị tối ưu của siêu tham số k là 3;

- Tại giá trị $k = 3$, MSE đạt giá trị rất nhỏ, và hệ số xác định của mô hình (R^2) xấp xỉ bằng 1, sai số phần trăm tuyệt đối trung bình (MAPE) bằng 0,31%. Ở đây ta cũng phân biệt sai số MSE trong quá trình xác thực trên bộ dữ liệu xác thực (Validation MSE) và sai số trong quá trình kiểm tra trên bộ dữ liệu kiểm tra (Test MSE). Trong đó, Validation MSE là tham số tối ưu, còn

Model Summary: K-Nearest Neighbors Regression

Nearest neighbors	Weights	Distance	n(Train)	n(Validation)	n(Test)	Validation MSE	Test MSE
3	rectangular	Euclidean	543	136	290	0.0000037954	0.0000017041

Note. The model is optimized with respect to the validation set mean squared error.



Hình 8. Kết quả xây dựng mô hình hồi quy dự báo hao tổn thủy lực tại phanh B1

Test MSE là thước đo khách quan sai số của mô hình.

Tương tự, các mô hình với độ chính xác rất cao (R^2 xấp xỉ bằng 1) cũng đã được xây dựng để xác định mất mát thủy lực tại phanh B2 và ly hợp C3, hao tổn cơ khí tại ly hợp C2, C3 và tại biển mô.

Ngoài ra, với các vị trí hư hỏng còn lại cũng xây dựng được các mô hình hồi quy dùng thuật toán KNN cho độ chính xác tương đối cao:

- Mất mát thủy lực tại ly hợp C1 với $R^2 = 0,772$;
- Mất mát thủy lực tại ly hợp C2 với $R^2 = 0,778$;
- Hao tổn cơ khí tại phanh B2 với $R^2 = 0,840$;
- Hao tổn cơ khí tại ly hợp C1 với $R^2 = 0,781$;

Để so sánh với các mô hình hồi quy truyền thống, nhóm tác giả xây dựng mô hình hồi quy tuyến tính cho 4 trường hợp có độ chính xác tương đối cao ở trên:

$$L = a + b_1 \cdot t_{100} + b_2 \cdot v_{30} + b_3 \cdot P_{30} \quad (1)$$

Trong đó, L là tổn hao cơ khí hoặc mất mát thủy lực (%), a là hệ số tự do, b_1 , b_2 và b_3 là các hệ số hồi quy tìm được bằng phương pháp tổng bình phương nhỏ nhất trên bộ dữ liệu đầu vào.

Kết quả cho thấy các mô hình xây dựng bằng thuật toán KNN có độ chính xác vượt trội, cụ thể là với các mô hình hồi quy tuyến tính truyền thống:

- Mất mát thủy lực tại C1 có $R^2 = 0,702 < 0,772$;
- Mất mát thủy lực tại C2 với $R^2 = 0,48 < 0,778$;
- Hao tổn cơ khí tại B2 với $R^2 = 0,739 < 0,840$;
- Hao tổn cơ khí tại C1 với $R^2 = 0,762 < 0,781$.

Kết quả xây dựng mô hình hồi quy tuyến tính xác định mất mát thủy lực tại ly hợp C1 trong 2 trường hợp đều cho kết quả là độ chính xác mô hình (R^2) thấp hơn so với mô hình được xây dựng bằng thuật toán

Bảng 1. Kết quả xây dựng mô hình hồi quy tuyến tính dự báo mất mát thủy lực tại ly hợp C1

Yếu tố	Hệ số hồi quy	p-value
Hệ số tự do	-4,294	<0,001
t_{100}	0,099	<0,001
v_{30}	0,108	<0,001
P_{30}	-0,031	0,059

$$R^2 = 0,703 \text{ (p-value} < 0,001)$$

Nguồn: Tác giả

Bảng 2. Kết quả xây dựng mô hình hồi quy tuyến tính dự báo mất mát thủy lực tại ly hợp C1 với mức ý nghĩa 0,05

Yếu tố	Hệ số hồi quy	p-value
Hệ số tự do	-4,999	<0,001
t_{100}	0,099	<0,001
v_{30}	0,083	<0,001

$$R^2 = 0,702 \text{ (p-value} < 0,001)$$

Nguồn: Tác giả

KNN. Ở trường hợp thứ nhất mô hình hồi quy tuyến tính truyền thống sử dụng phương pháp tổng bình phương nhỏ nhất và chấp nhận tất cả các yếu tố đầu vào (không sử dụng mức ý nghĩa p-value) để tối đa giá trị hệ số xác định, ta thu được $R^2 = 0,703$ (Bảng 1). Phương trình hồi quy có dạng như sau:

$$L = -4,294 + 0,099 \cdot t_{100} + 0,108 \cdot v_{30} - 0,031 \cdot P_{30};$$

$$R^2 = 0,703 \quad (2)$$

Trong đó, biến đầu vào $P30$ có mức ý nghĩa ($p\text{-value} = 0,059$) khi kiểm định bằng bài kiểm định t-test. Trên cơ sở đó, ở trường hợp thứ hai, mô hình hồi quy tuyến tính với mức ý nghĩa $p\text{-value} < 0,05$ (loại bỏ biến $P30$) cho hệ số xác định $R^2 = 0,702$ (Bảng 2). Phương trình hồi quy có dạng như sau:

$$L = -4,999 + 0,099.t_{100} + 0,083.v_{30};$$

$$R^2 = 0,702 \quad (3)$$

Qua so sánh các giá trị R^2 ở hai trường hợp cho thấy việc loại bỏ biến $P30$ không làm giảm chất lượng mô hình hồi quy tuyến tính một cách đáng kể.

4. Kết luận

Nghiên cứu đã mô phỏng các dạng hư hỏng thủy lực và cơ khí trong hộp số tự động của xe ô tô trong điều kiện vận hành tại Việt Nam, sử dụng ba đặc trưng đầu vào ($t100$, $v30$, $P30$) để xây dựng mô hình KNN phân loại với độ chính xác đạt 92,2%. Phân tích hoán vị đặc trưng cho thấy $t100$ có ảnh hưởng lớn nhất, trong khi $v30$ và $P30$ tuy tương quan cao vẫn đóng vai trò cần thiết. Dựa trên dữ liệu này, 11 mô hình hồi quy sử dụng thuật toán KNN đã được phát triển để dự báo hao tổn cơ khí hoặc mất mát thủy lực tại vị trí hư hỏng; trong đó 7 mô hình đạt R^2 xấp xỉ 1 và 4 mô hình còn lại đạt R^2 từ 0,77 - 0,84, cho thấy hiệu quả vượt trội so với hồi quy tuyến tính truyền thống.

Các hướng nghiên cứu tiếp theo bao gồm: (i) kiểm chứng các mô hình đã xây dựng trong nghiên cứu này bằng các kết quả thực nghiệm để nâng cao độ tin cậy; (ii) so sánh KNN với các thuật toán học máy khác như (Support Vector Machine) SVM, Random Forest, Gradient Boosting cũng như các thuật toán học sâu (Deep Learning) để lựa chọn giải pháp tối ưu. Kết quả nghiên cứu cũng gợi mở khả năng phát triển module phần mềm giám sát thời gian thực, tích hợp vào hệ thống điện tử của xe để theo dõi tình trạng kỹ thuật hộp số tự động, cảnh báo sớm cho người lái và hỗ trợ công tác bảo dưỡng, sửa chữa. Ngoài ra, cơ sở dữ liệu và mô hình đề xuất có thể được sử dụng trong thiết kế giả thử nghiệm phục vụ đánh giá tình trạng kỹ thuật hộp số trước và sau sửa chữa.

TÀI LIỆU THAM KHẢO

- [1] Gama, Renann & Santos, Ramon & Cabette, Regina & Prado, Pedro & Cunha, Kathleen & Silva, Carlos. (2023). *Study And Analysis of Fractures in Gears from Automotive Transmissions*. 10.13140/RG.2.2.27444.73609.
- [2] Chen, Li-dan & Chen, Kai-cao. (2012). *The Application of Atomic Absorption Spectrometry in Automatic Transmission Fault Detection*. Guang pu xue yu guang pu fen xi = Guang pu. 32. 244-7. 10.3964/j.issn.1000-0593(2012)01-0244-04.
- [3] Merksiz, Jerzy & Pielecha, Jacek & Pacholek, Adam. (2008). *Directions of on Board Diagnostic System Development in Passenger Cars*. Journal of Konbin. Vol.4. pp.165-186. 10.2478/v10040-008-0017-7.
- [4] S, Avinash & Ramachandra, & Prathapchandra,. (2025). *Advanced Transport Management System (aTraMS) using On-Board Diagnostics*. 10.21203/rs.3.rs-5899243/v1.
- [5] Zhang, Hongxin & Ji, Zhongshu & Song, Ruijin. (2024). *Research on Typical Fault Diagnosis of Gear based on GHM Multiwavelet Transform and MCKD Algorithm*. Recent Patents on Engineering. Vol.18. 10.2174/1872212118666230721153234.
- [6] Zuñiga, Mariana & Rojas, Rafael & Nápoles, Elías & Martínez, Moisés. (2024). *Non-invasive analysis for detecting gear fractures in automotive transmissions*. Journal of Scientific and Technical Applications. 10.35429/JSTA.2024.10.25.1.7.
- [7] Su, Chengyun & Liu, Dunning & Zhao, Huanyu & Yuan, Yanling & Song, Tingbin. (2024). *Correlation of the sound quality and vibration of end of line testing for automatic transmission*. Measurement and Control. 57. 1405-1416. 10.1177/00202940241241217.
- [8] S, Aditya & Venkatesh S, Naveen & Chakrapani, Ganjikunta & Sugumaran, V.. (2024). *Deep learning based fault detection of automobile dry clutch system using spectrogram plots*. Engineering Research Express. Vol.6. 10.1088/2631-8695/ad4771.
- [9] Yuan, Haibing & Yang, Yiyang & Guo, Bingqing & Zhao, Fengsheng & Zhang, Di & Yang, Shuai. (2024). *Research on gear flank surface defect detection of automotive transmissions based on improved YOLOv8*. Measurement Science and Technology. 35. 10.1088/1361-6501/ad748b.
- [10] Badiger, Aishwaryalaxmi & Priyadarshi, Priyamvad & Bhat, Goutam. (2024). *Dynamic Automatic Transmission Oil Life Prediction Using Machine Learning*. 10.4271/2024-26-0036.
- [11] Alshammari, Ahmad. (2024). *Implementation of*

Classification using K-Nearest Neighbors (KNN) in Python. International Journal of Computer Applications. Vol.186. pp.19-24. 10.5120/ijca2024923894.

[12] Nearest Neighbors - scikit-learn [Online], Feb 21 2025. Available: <https://scikit-learn.org/stable/modules/neighbors.html>

Ngày nhận bài:	12/05/2025
Ngày nhận bản sửa:	03/10/2025
Ngày duyệt đăng:	17/10/2025