วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2025.12.015

บทความวิจัย

A Comparative Analysis of Supervised Machine Learning Algorithms for Fault Prediction in Automotive Suspension Systems

Pachara Juyploy[1*] and Withit Chatlatanagulchai[2]

[1] Materials Handling and Logistics Engineering, Faculty of Engineering, King Mongkut's University of Technology North Bangkok

[2] Mechanical engineering, Faculty of Engineering, Kasetsart University

[*] Corresponding author, E-mail: pachara.j@eng.kmutnb.ac.th

Received: 19 July 2024; Revised 18 November 2025; Accepted: 8 December 2025
Online Published: 12 December 2025

Abstract: Intense automotive vibrations, while common, can seriously compromise driver health. This research utilizes machine learning (ML) to predict potential failures in car suspension systems, targeting an enhancement in vehicle reliability and safety. While numerous studies have simulated suspension faults, the increasing data complexity from uncertain parameters necessitates more efficient algorithms for precise fault identification. This study, therefore, conducts a comparative analysis of several supervised machine learning algorithms to determine the most accurate method for this predictive task. The algorithms were evaluated using four distinct feature set preparations: original data, standard deviation data, principal component analysis data, and a combined set of mean standard deviation and principal component analysis. The findings reveal that the Artificial Neural Network (ANN) and Support Vector Classifiers (SVC) algorithms yield the highest prediction accuracy. Notably, this peak accuracy was achieved when utilizing the combined feature set (mean standard deviation and Principal Component Analysis (PCA)). These results offer a valuable contribution toward designing more robust car suspension systems and advancing future preventive maintenance strategies.

Keywords: Suspension Systems; Machine Learning Algorithm; Predictive Maintenance; Supervised Learning Techniques

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2025.12.015

1. Introduction

The impact of whole-body vehicle vibrations on human health has been a subject of significant investigation. Occupational health data, for example, highlights accidents linked to driver fatigue following prolonged vibration exposure [1]. This exposure is associated with numerous health risks, including chronic back pain, digestive disorders, and hearing damage [2]. Consequently, research has focused on vehicle suspension system enhancements, such as active controls to optimize ride smoothness [3] and studies on the vibration impacts of varying road conditions [4]. To better address these challenges, machine learning (ML) is increasingly being adopted as a powerful tool for predicting and assessing potentially harmful variations in suspension systems.

The predictive capabilities of ML have been proven effective across diverse industries, such as forecasting brake quality defects [5] and predicting critical health outcomes in clinical settings [6, 7]. Within vehicle diagnostics, ML applications are varied [8]. While some studies have focused on fault detection using unsupervised clustering methods [9-12], the use of supervised techniques is particularly promising for fault prediction. These algorithms ranging from simpler statistical models like Naïve Bayes (NB) to complex, non-linear models like Artificial Neural Networks (ANN) and Support Vector Classifiers (SVC) are selected for this study because they represent a broad and well-established spectrum of supervised learning approaches, allowing for a robust comparison of their suitability for this specific diagnostic task.

Despite this growing body of research, a significant research gap exists. There is a lack of comprehensive, direct comparison studies that evaluate a wide range of supervised ML algorithms using the same standardized dataset for car suspension fault prediction. Many existing studies focus on only one or two algorithms, making it difficult for engineers to determine which method is truly optimal for this application.

Therefore, this study aims to fill this gap by conducting a robust comparative analysis of eight widely used supervised and unsupervised ML techniques. The selected algorithms include one unsupervised technique, K-Means (KM) Clustering, and seven supervised techniques: K-Nearest Neighbors (KNN), Support Vector Classifiers (SVC), Logistic Regression (LR), Linear Discriminant Analysis (LDA), Naïve Bayes (NB), Artificial Neural Network (ANN), and Decision Tree (DT). This research systematically validates the accuracy of these algorithms, employing careful parameter tuning to prevent underfitting and overfitting [13, 14].

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2025.12.015

This paper provides a detailed evaluation, highlighting how specific preprocessing strategies, combined with optimal algorithm tuning, contribute to achieving the highest classification accuracy for car suspension failures.

2. Materials and Methods

To accurately diagnose faults within the automotive suspension system, this study implements a dual-stage machine learning approach. Initially, an unsupervised K-means clustering technique is applied to discover inherent grouping structures within the vibration data. Subsequently, a comprehensive suite of seven supervised learning algorithms specifically KNN, SVC, LR, LDA, NB, ANN, and DT is utilized to construct predictive models capable of classifying the specific degree of suspension damage. The car suspension vibration data is divided into two main parts: the training dataset and the validation dataset. The training dataset is used to calibrate the algorithm's parameters, while the validation dataset is used to evaluate the algorithm's performance [15]. The process began by assessing the unsupervised learning method using K-means, which involves determining the specified number of clusters and using the Calinski-Harabasz Criterion to find the appropriate number of clusters. Once the exact number of clusters is determined, supervised learning methods are used to predict car suspension damage.

2.1 Datasets

The acceleration data obtained from the left front wheel suspension is studied, and the car parameters are referenced from [16]. The event occurs at the 20-second mark, with the spring coefficient reductions starting from 10% up to 100% in 2% increments, at a sampling frequency of 100 Hz. At each increment of spring reduction, the simulation generates 2000 data points, corresponding to a duration of 20 seconds. The fault scenario in the spring coefficient ranges from 60% to 100% reduction. The Data scenarios are categorized into four levels: Good, Usual, Fail, and Fail High, as shown in Table 1.

Table1 Data scenarios

Categories	Spring Coefficient Reduction (%)
Good	10% - <35%
Usual	45% - <65%
Fail	65% - 80%
Fail high (Severe)	90%-100%

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2025.12.015

2.2 Preparing for Datasets

Dataset preparation involved extracting raw acceleration signals from the car suspension damage simulation. This data capture encompassed a full spectrum of vibration patterns, ranging from normal operation to severe irregularity. Unsupervised machine learning was then applied to this raw data to mathematically determine the optimal number of damage clusters.

To ensure robust model evaluation, the collected data was partitioned using an 80:20 ratio, allocating the majority for model training and the remaining 20% for validation purposes. The training subset served to calibrate the supervised algorithms, whereas the validation subset facilitated performance assessment. Furthermore, to optimize model hyperparameters and verify accuracy stability, a 5-fold cross-validation strategy was implemented. Statistical enhancement of the dataset was achieved through two primary preprocessing methods: Standard Deviation normalization and Principal Component Analysis (PCA).

To rigorously test algorithmic performance under different conditions, four distinct feature sets were derived from the primary data: 1) The unprocessed 'Raw' signals; 2) Data standardized via Standard Deviation; 3) A dimensionality-reduced set using PCA (retaining 80% variance); and 4) A comprehensive set combining Mean, Standard Scaler, and PCA. These variations were specifically engineered to identify which preprocessing strategy yields the highest diagnostic accuracy.

2.3 Tools

MATLAB was used for initial data preparation, including standard deviation and PCA transformation of the data. All unsupervised and supervised algorithms were executed using MATLAB Live Editor and MATLAB Classification Learner [17], including generating all result graphs

3. Clustering Unsupervised ML

K-means clustering [18], [19] was selected to partition the suspension vibration data into distinct fault categories without requiring pre-labeled training examples. The optimal number of clusters for this specific dataset was identified as K=4, derived from the Calinski-Harabasz Criterion analysis (Fig. 1) [20]. To ensure robust grouping based on feature similarity, the algorithm utilized the Squared Euclidean distance metric. Configuration settings included a single replicate run with a maximum of 100 iterations, ensuring convergence on the complex vibration patterns. Each cluster represents distinct patterns of acceleration measurements associated with specific levels of suspension damage. The algorithm discerns various types and levels of damage, detailed in Table 2 and Fig. 2. The data preprocessing involved mean, standard deviation, and PCA techniques

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2025.12.015

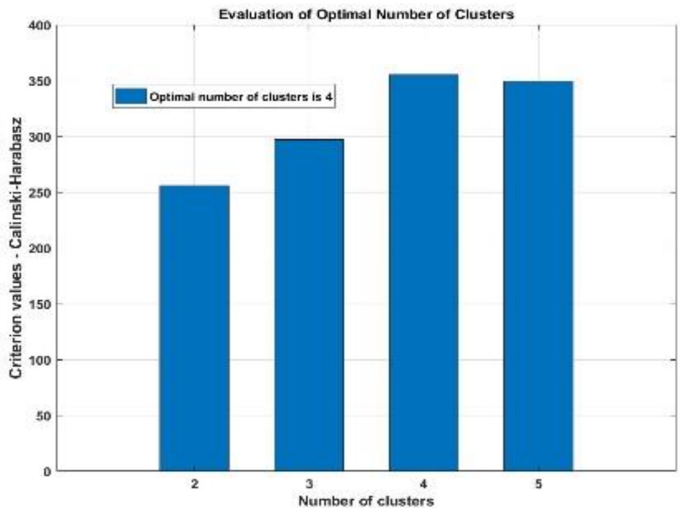

Fig. 1 optimal K=4 using the Calinski-Harabasz Criterion

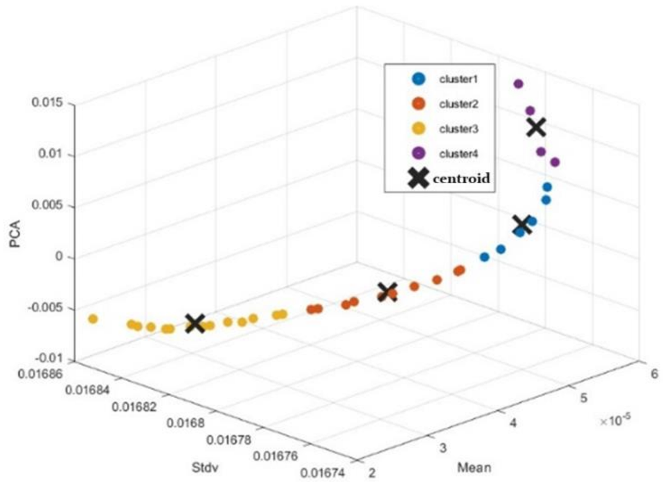

Fig. 2 Four clusters acceleration data using the K-Means Clustering

Table 2 shows 4 divisions of data data clusters when examine the springs reduction grouping by K-Means Clustering. It is consistent with the data scenarios defined in the car suspension damage simulation.

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2025.12.015

Table 2 Spring reduction clusters acceleration data using the K-Means Clustering

Spring reduction	Cluster	Membership %	Centroid points Mean ($x10^{-5}$)	Stdv ($x10^{-2}$)	PCA ($x10^{-3}$)
0% - 40%	3	41.1	2.65x	1.682	-5.28
43% - 65%	2	29.4	3.62	1.677	-40.38
68% - 80%	1	17.6	4.69	1.675	5.654
83% - 90%	4	11.7	5.66	1.677	11.00

Table 3 Tuning parameters

Tuning Parameters	Detail Description
Ne	Number of neighbors
B	Box constraint level
K	Kernel scale
N F S	Number of layers First Layer size Second Layer size
Sp	Maximum Number of splits

Table 4 Types of the dataset

Types of the Dataset	Detail Description
Raw	Original data
St	Standard deviation
PCA	Principal Component Analysis
M **+** St**+** PCA	Mean + Standard deviation + PCA

4. Predicting using Supervised ML

Parameters are employed to optimize the algorithms prior to applying supervised machine learning. Table 3 defines five distinct sets of parameters for five different algorithms, aimed at identifying optimal performance across four dataset types (Raw, Standard Deviation, PCA, and Mean + Standard Deviation + PCA). The best hyperparameters were determined through various parameter configurations to assess the suitability of each dataset. Table 4 Show types of datasets.

Algorithmic tuning for the KNN, SVC, LR, and ANN models was executed based on the specific configurations detailed in Table 3. Conversely, the LDA and NB classifiers were deployed using their standard default settings. Comprehensive training and testing procedures were applied to all algorithms across the four prepared dataset variants. The resulting performance metrics for each model are presented in Table 5.

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2025.12.015

Table 5 Accuracy for each algorithm

Algorithm	Types of the dataset	Tuning parameters	Accuracy (%)
KNN	Raw	Ne=3	68.5
	St	Ne=11	54.8
	PCA	Ne=7	91.8
	M+ St + PCA	Ne=1	95.9
SVC	Raw	B=5	74.0
	St	B=1	50.7
	PCA	B=1	91.8
	M+ St + PCA	B=4	98.6
LR	Raw	K=4	53.1
	St	K=5	31.5
	PCA	K=1	31
	M +St + PCA	K=1	82.2
LDA	Raw	Default	94.5
	St	Default	42.5
	PCA	Default	80.6
	M +S t+ PCA	Default	83.8
NB	Raw	Default	78.1
	St	Default	56.2
	PCA	Default	86.3
	M +St + PCA	Default	93.2
ANN	Raw	N=2 F=10 S=10	90.4
	St	N =1 F=10	52.1
	PCA	N =2 F=10 S =10	94.5
	M +St + PCA	N=1 F=10	98.6
DT	Raw	Sp = 100	95
	St	Sp =100	43.8
	PCA	Sp = 100	87.7
	M +St + PCA	Sp = 100	89.0

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2025.12.015

The training and comparison of all machine learning models were conducted using the MATLAB Classification Learner App. For complex models (KNN, SVC, ANN, and DT), the app's automated hyperparameter tuning function was utilized to find the optimal parameter values that yielded the highest validation accuracy. For simpler models (LR, LDA, and NB), the app's default parameter settings were employed as they generally do not require extensive tuning for this application.

The test results proved that various algorithms perform better when data is preprocessing, and the parameters are tuned. The results showed that KNN, SVC, LR, NB, and ANN achieved the highest accuracy when using preprocessing with Mean + Standard Deviation + PCA. However, LDA and DT performed better with raw data and slightly had lower accuracy when using preprocessing with Mean + Standard Deviation + PCA. In conclusion, Mean + Standard deviation + PCA leads to optimal accuracy results. Furthermore, fine-tuning parameters and scaling data can enhance algorithmic precision.

The summary of the top three algorithms achieving the highest accuracy, as presented in Table 6, identifies the best-performing algorithms in the following order: ANN, SVC, and KNN. These algorithms achieve near-perfect accuracy, utilizing datasets that have been preprocessed with Mean + Standard deviation + PCA to attain optimal performance.

Table 6 The best algorithms

Type of the datasets	Algorithm	% Accuracy
M+St+ PCA	ANN	98.6
M+St+ PCA	SVC	98.6
M+St+ PCA	KNN	95.9

4. Discussion and Conclusion

Although car vibrations are typically considered normal, it's crucial to monitor their intensity because excessive vibrations can negatively affect human health. Moreover, unsupervised and supervised machine learning algorithms are powerful techniques for clustering, categorizing damage, and predicting uncertain and complex vehicle damage data.

For example, a related study by Saleh and Fleyeh (2022) [21] also utilized supervised machine learning to predict the status of road signs. Their study investigated the effects of using principal component analysis and data scaling on prediction accuracy. The authors of that paper applied three algorithms: Random Forest, Artificial Neural Network (ANN), and

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2025.12.015

Support Vector Machines (SVM). In their findings, Random Forest achieved the highest accuracy, precision, recall, and F1 score, all at 98%.

In contrast, the findings from the present study, which focused on car suspensions, revealed that the ANN and SVC algorithms achieve the highest accuracy of 98.6% when using the dataset of the combined mean standard deviation and principal component analysis data.

This superior performance (from ANN and SVC) can be attributed to the inherent ability of both algorithms to manage the complex, non-linear relationships present in the vehicle suspension vibration data.

ANN (Artificial Neural Network): Is designed to learn intricate patterns through its multi-layer structure. This made it highly effective at detecting the subtle, complex relationships within the feature-engineered dataset (Mean + Stdv + PCA) that other models missed.

SVC (Support Vector Classifiers): Excels in high-dimensional spaces. By utilizing the "Kernel Trick" (which the MATLAB app employs), SVC can create a flexible, non-linear decision boundary (hyperplane). This allowed it to effectively separate the complex 'Fault' and 'Good' classes, a task where simpler linear models failed.

Conversely, models like Logistic Regression (LR) and Linear Discriminant Analysis (LDA) performed poorly because they are limited by their linear assumptions, which were not suitable for the complexity of this dataset. Similarly, Naïve Bayes (NB) was hindered by its 'naïve' assumption that all features are independent, which is not true for highly correlated vibration sensor data.

This analysis underscores that for complex fault prediction tasks involving vibration data, non-linear models like ANN and SVC are the most robust and appropriate choices.

4.1 Practical Implications

The high accuracy of ANN and SVC models enables their use in an on-board diagnostic (OBD) system. This system could process real-time vibration data, alert the driver to potential faults, and transform maintenance from reactive to predictive. This directly enhances driver safety, reduces long-term repair costs, and increases vehicle reliability.

4.2 Limitations and Future Work

The primary limitation is the use of simulation data, which may not capture real-world complexities; the study also focused only on spring faults. Future work must validate these models using data from real vehicles and expand the system to detect other fault types (e.g., damper failure). Investigating deep learning models (like CNNs or RNNs) could also yield further improvements.

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2025.12.015

5. Acknowledgement

This work was supported by the Faculty of Engineering under the funding agreement No. ENG-NEW-63-30. The authors would like to express their sincere gratitude for this valuable support, which made the completion of this study possible.

6. References

[1] Land Transportation Safety and Health Association, Explanation of safety guidelines for cargo handling operations. 2018. http://rikusai.or.jp/wp-content/uploads/pdfs/niyaku-guideline_kaisetsu_H30.pdf (Accessed on 22 August 2022) (in Japanese)

[2] J. Tatsuno and S. Maeda, Effect of whole-body vibration exposure in vehicles on static standing balance after riding, Vibration, 2023, 6(2), 343–358.

[3] S.S. Sedeh, R.S. Sedeh and K. Navi, Reducing harmful effects of road excitations on human health by designing car active Suspension systems, International Conference on Bioinformatics and Computational Biology (BIOCOMP'06), Proceeding, 2006, 236-241.

[4] U. Kırbas, Investigation of the effects of whole-body vibration exposure on vehicle drivers when travelling over covered manholes embedded in public roadways, International Journal of Industrial Ergonomics, 2022, 88, 103277.

[5] MH. Yorulmus, HB. Bolat and C. Bahadird, Predictive quality defect detection using machine learning algorithms: A case study from automobile industry, InIntelligent and Fuzzy Techniques for Emerging Conditions and Digital Transformation INFUS 2021 Conference, Proceeding, 2022, 263-270.

[6] G. Kong, K. Lin and Y. Hu, Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU, BMC Medical Informatics and Decision Making, 2020, 20, 251.

[7] V. Ganesan M.P. Talawar, H. Pant, M.S. Sayyed, M.S. Tamboli, M.N. Shaik and M.N. Shaik, Predictive maintenance in healthcare IoT: A machine learning-based approach, European Chemical Bulletin, 2023, 12(4), 15909-15922.

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2025.12.015

[8] A. Theissler, J. Perez-Velazquez, M. Kettelgerdes and G. Elger, Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry, Reliability Engineering and System Safety, 2021, 215, 107864.

[9] G. Wang and S. Yin, Data-driven fault diagnosis for an automobile suspension system by using a clustering based method, Journal of the Franklin Institute, 2014, 351(6), 3231-3244.

[10] S. Yin and Z. Huang, Performance monitoring for vehicle suspension system via fuzzy positivistic C-means clustering based on accelerometer measurements, IEEE/ASME Transactions on Mechatronics, 2014, 20(5), 2613-2620.

[11] Y. Wang, X. Zheng, L. Wang, G. Lu, Y. Jia, K. Li and M. Li, Sensor fault detection of vehicle suspension systems based on transmissibility operators and Neyman–Pearson test, Reliability Engineering and System Safety, 2023, 232, 109067.

[12] K. Wu, J. Liu, M. Li, J. Liu and Y. Wang, Multi-mode active suspension control based on a genetic K-means clustering linear quadratic algorithm, Applied Sciences, 2021, 11(21), 10493.

[13] R. Tabares-Soto, S. Orozco-Arias, V. Romero-Cano, V.S. Bucheli, J.L. Rodríguez-Sotelo J, and C.F. Jiménez-Varón, A comparative study of machine learning and deep learning algorithms to classify cancer types based on microarray gene expression data, PeerJ Computer Science, 2020, 6, e270.

[14 J. Zou, M. Huss, A. Abid, P. Mohammadi, A. Torkamani, and A. Telenti, A primer on deep learning in genomics , Nature Genetics, 2018, 51(1),12-18.

[15] G. Eraslan, Z. Avsec, J. Gagneur, and FJ. Theis, Deep learning: New computational modelling techniques for genomics, Nature Reviews Genetics, 2019, 20 (7), 389–403.

[16] P. Juyploy and W. Chatlatanagulchai, Reducing harmful effects on human health of car vibrations using a combination of unsupervised and supervised machine learning algorithm, The Journal of Industrial Technology, 2024, 20 (1), 171–185

[17] S.M. Senthamil, and S.J. Rani, Classification of admission data using classification learner toolbox, In Journal of Physics: Conference Series, 2021, 1979, 012043.

วารสารวิชาการเทคโนโลยีอุตสาหกรรม (The Journal of Industrial Technology)
ISSN (online): 2697-5548
DOI: 10.14416/j.ind.tech.2025.12.015

[18] S. Eke, T. Aka-Ngnui, G. Clerc and I. Fofana, Characterization of the operating periods of a power transformer by clustering the dissolved gas data, IEEE 11th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED), Proceeding, 2017, 298-303.

[19] N.L. Wibowo, M.A. Soeleman and A.Z. Fanani, Antlion optimizer algorithm modification for initial centroid determination in K-means algorithm, Jurnal RESTI, 2023, 7(4), 870-883.

[20] S. Lukasik, P.A. Kowalski, M. Charytanowicz, and P. Kulczycki, Clustering using flower pollination algorithm and Calinski-Harabasz index, IEEE Congress on Evolutionary Computation (CEC), Proceeding, 2016, 24, 2724-2728.

[21] R. Saleh and H. Fleyeh, Using supervised machine learning to predict the status of road signs, Transportation Research Procedia, 2020, 62, 221-228.