Fault Diagnosis in Vehicle Air Conditioning Systems via Comparison of k-Nearest Neighbors and Random Forest Classification Models

Dinh Anh Tuan Tran*, Thi Khanh Phuong Ho, and Van Tuan Nguyen

Faculty of Heat & Refrigeration Engineering, Industrial University of Ho Chi Minh City, Ho Chi Minh 700000, Vietnam

Email: trandinhanhtuan@iuh.edu.vn (D.A.T.T); hothikhanhphuong@iuh.edu.vn (T.K.P.H); nguyenvantuan@iuh.edu.vn (V.T.N)

*Corresponding author

***Abstract*—Fault Detection and Diagnosis (FDD) of vehicle air conditioning (A/C) system is always a vital technique for achieving energy-saving goals and maintaining system reliability. The performance of k-Nearest Neighbors (kNN) and Random Forest (RF) models are invested for diagnosing faults in vehicle A/C systems. Two frequent faults, condenser fouling and refrigerant leakage, are selected in this study. A total of 745 validation samples and 568 test samples, covering normal operation and seven fault conditions with varying levels of these two faults, were analyzed. Model performance was evaluated using accuracy, precision, recall, F1-Score, and Receiver Operating Characteristics (ROC) curve metrics. Additionally, the confusion matrix was employed to provide a detailed breakdown of a model's classification performance. Results showed that both models achieved high validation accuracy (~91.68%), with RF slightly outperforming kNN in testing (RF: 90.26%). The kNN model exhibited higher recall, enhancing the detection of true positive faults, whereas RF demonstrated better balance between precision and F1-Scores. ROC curve analysis further confirmed that RF provided better discrimination of overlapping fault classes. Confusion matrix results indicated that both models struggled with intermediate levels of condenser fouling, revealing a need for improved fault differentiation. Overall, the RF model demonstrated greater robustness and consistency, making it more suitable for reliable diagnosis of vehicle A/C faults.**

***Keywords*—fault, diagnosis, machine learning model, k-nearest neighbor, random forest**

I. Introduction

The energy conversation has been a topic of concern for decades. End-use sectors, including the transportation, industrial, residential, and commercial sectors, consume large amounts of energy. Among these, according to Ağbulut [1], transportation accounts for 28% of the total energy consumption. Meanwhile, vehicles are still an indisputable element of the human experience. Currently, the demand for vehicles is increasing, yet global fuel supplies remain limited. As a result, past studies have consistently shown an interest in the energy conservation of vehicles. However, several factors can affect a car's fuel consumption, such as mechanical failure, driving competence, and the Air Conditioning (A/C) system. Among these factors, the A/C system, which is crucial for modern vehicles owing to its ability to provide thermal comfort, consumes the most energy: it requires up to 30% of the total fuel use or more than 27 billion liters of gasoline annually in the United States [2].

Consequently, reducing energy usage has become a pressing requirement of vehicle A/C systems. Researchers have conducted various studies to reduce the energy consumption of automotive A/C systems. Generally, researchers found a close relationship between energy waste and A/C unit failure performance. According to several studies [3, 4], when the A/C system is in a fault state, it wastes about 30%–40% of the total energy consumed [5]. If the operating state of the A/C system can be regulated to ensure that it always functions under normal conditions, energy savings can account for 40% of the total energy usage. Therefore, a robust method for detecting and diagnosing faults in vehicle A/C systems is required to achieve this purpose.

Machine learning has been widely researched and applied to cars to assist humans, such as in identifying traffic signals, pedestrians, road marking, and ensuring distances [6–8]. Thanks to machine learning, such vehicles can increase accessibility, improve road safety, and reduce accidents. According to a study published [9], employing machine learning techniques in vehicles has

Manuscript received February 7, 2025; revised March 31, 2025; accepted May 22, 2025; published October 14, 2025.

doi: 10.18178/ijmerr.14.5.534-541

benefits over not using machine learning. In addition, machine learning has been developed in the fields of fault detection and diagnosis in A/C systems. Tran *et al.* [10] employed the RBF method to detect and diagnose faults in chiller systems. The results showed that the RBF method was accurate for the detection and diagnosis of six chiller faults. Refrigerant leak detection based on machine learning, was reported by Lei *et al.* [11]. The results of refrigerant leak detection in vehicles showed that the accuracy of the proposed method can reach 95.69%. Similarly, the development of machine learning for A/C fault detection and diagnosis has also been discussed in various published works such as support vector regression [12, 13], the Kriging method [14], and artificial neural networks [15, 16]. In general, machine learning method methods are widely applied in fault detection and diagnosis in commercial and industrial A/C systems. Conversely, limited publications have addressed its application to A/C systems in vehicles. This study investigates two classification models, the k-Nearest Neighbor (kNN) algorithm and the Random Forest (RF), as the A/C fault detection and diagnosis model. Next, the A/C fault detection and diagnosis Fault Detection and Diagnosis (FDD) proposed a useful technique to improve the energy-saving goals for two frequent faults.

II. Research Methodology

A. The kNN Algorithm for Classification Model

The kNN method is a commonly used in machine learning method. Scholars commonly use it for classification because of its simplicity and high accuracy. The kNN technique is based on a given data matrix with M rows and N columns. The prediction of a query point is based on the similarity of the data points in a given dataset. Generally, the kNN algorithm finds the K-nearest neighbors to a given data point using a distance metric between the query and other points in the data matrix. Here, K represents the number of nearest neighbors to be considered when making predictions. The distance metric can employ one of the following distance functions: Euclidean distance (Eq. (1)), Manhattan distance (Eq. (2)), or Minkowski distance (Eq. (3)). The weighted or majority vote of the K neighbors then determines the class or value of the query data point. Therefore, the system can easily identify the K-nearest points to the query by sorting the distances in ascending order and retrieving the K points that have the shortest distance between the dataset and the query. Typically, the best way to select the best value for K is to first inspect the data. Although reducing the total noise makes the higher K values more precise, this is not a guarantee. Cross validation is another common method to find an acceptable K value between 3 and 10.

$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{1}$$

$$\sum_{i=1}^{k}|x_i - y_i| \tag{2}$$

$$\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{\frac{1}{q}} \tag{3}$$

where x_i and y_i are the query point *ith* and a case from the training data sample, respectively.

Two well-known voting method in the kNN classification are presented in the following equations.

Majority voting:

$$y' = arg\ max_y \sum_{(x_i, y_i) \in D_K} \delta(y, y_i) \tag{4}$$

Distance-Weighted Voting:

$$y' = arg\ max_y \sum_{(x_i, y_i) \in D_K} w_i \delta(y, y_i) \tag{5}$$

where:
- y' is the predicted label for the test point data.
- D_K is the data set of the kNN of the test sample; x_i and y_i denotes the data and the class label in D_K, respectively. Cross validation was used to select the K value.
- $\delta(y, y_i)$ = 1 if $y_i = y$ and 0 otherwise.
- w_i is the weighted distance of x_i which is determined by the Euclidean distance metric as follows:

$$w_i = exp\left(\frac{-\|x - x_i\|_2^2}{a^2}\right) \tag{6}$$

where *a* is an optional positive number.

B. The Random Forest Machine Learning For Classification Model

RFs are effective machine learning models for prediction. Injecting appropriate randomness makes them accurate classifiers and regressors. Furthermore, the framework, which considers the strength of individual predictors and their correlations, provides insight into the predictive capabilities of RF. RFs consist of an ensemble of tree predictors, where each tree relies on the values of a randomly selected vector, which is independently drawn and identically distributed over all trees in the forest. The generalization error in forests approaches a limit when the number of trees in the forest increases significantly. Consider an ensemble of classifiers, $u_1(x), u_2(x), \ldots u_n(x)$ with the training set randomly sampled from the distribution of the random vectors *Y* and *X*. We define the margin function as follows:

$$mg(X, Y) = a\vartheta_n I(u_n(X) = Y) - max_{j \neq Y} a\vartheta_n I(u_n(X) = j) \tag{7}$$

where *X* represents the input measure, $a\vartheta_n$ denotes the average number of votes at *X* and *Y* for the respective classes, and $I(.)$ signifies the indicator function. The margin quantifies the degree to which the average vote count for classes *X* and *Y* exceeds the average vote for any alternative class. A larger margin indicates greater classification confidence. The generalization error is expressed as follows:

$$PE^* = P_{X,Y}(mg(X, Y)) < 0 \tag{8}$$

where the subscripts X and Y indicate that the probability lies within the X and Y space.

III. Faults Selection And Simulated Experimental Faults

A. Faults Selection

A fault is defined as an unpermitted deviation of at least one characteristic property of a variable from the acceptable behavior. A fault can result in inefficiency, malfunction, or even system damage. Therefore, an accurate fault detection tool for the A/C system in modern vehicles is crucial for maintaining the system's functionality and saving energy. The AC system is known to have numerous faults. Generally, these faults can be roughly distinguished into two major categories: abrupt and gradual faults. The obvious symptoms of abrupt faults such as locked compressors, burned magnetic compressor coils, and broken pipes, make them easy to address. Therefore, these faults are rarely the subject of scholars. Otherwise, common faults in the A/C system, such as refrigerant overcharge, refrigerant leakage, non-condensable gas, evaporator air blockage, and condenser air blockage, tend to occur gradually. These faults are typically difficult to detect because of their unclear symptoms. They only react when a feature undergoes a significant change, such as a large, sudden, or long-lasting, gradually increasing fault. Typically, they emerge when the system encounters significant operational problems; this implies that the indications of an A/C system failure are readily apparent. We selected refrigerant leakage and condenser fouling as these faults because their frequent occurrence in Vietnam.

The time dependency of the faults can be distinguished, as shown in Fig. 1, as an abrupt fault (stepwise), an incipient fault (drift-like), and an intermittent fault.

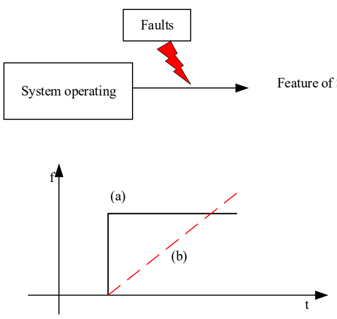


Fig. 1. Time-dependency of faults: (a) abrupt; (b) incipient.

B. Experimental Data And Method

In this study, refrigerant R134a was used in a conventional A/C system consisting of a belt-driven compressor, an air-cooled condenser, a laminar microchannel evaporator, and a thermostatic expansion valve. These components supported the development and evaluation of the proposed FDD methods, with the simulated faults detailed in Table I. Eight fault-related feature parameters were selected to reliably capture the system's health state. Four PT-100 temperature sensors, each with typical accuracy of ±0.15 °C, were positioned to monitor the refrigerant temperatures at the condenser's inlet (Tci) and outlet (Tco), and at the evaporator's inlet (TEI) and outlet (TEO). Additional sensors recorded air temperatures at the condenser's entry (Tai) and exit (Tao), enabling the calculation of the condensing (Tcd) and evaporating (Tev) saturation temperatures. The temperature sensors are arranged as shown in Fig. 2.

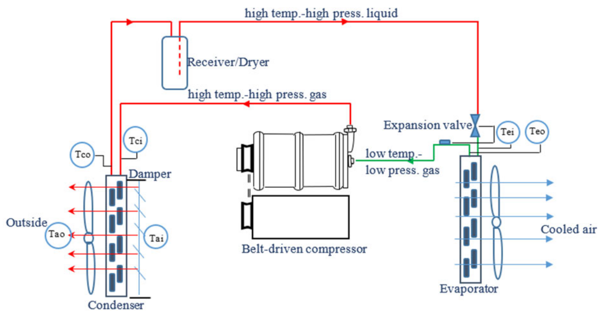


Fig. 2. Structure diagram of the test for vehicular A/C system.

Two types of experiments were conducted: fault-free and fault experiments. As mentioned above, two common faults, namely, refrigerant leakage and condenser fouling were chosen in the simulated experiment. Each fault was simulated at three distinct severity levels. For refrigerant leakage, the R134a refrigerant charge was incrementally reduced by 10% (525 g for normal). The amount of R134a refrigerant added was weighed on a commercial digital scale with a measurement error of ±5 g. For condenser fouling, the airflow through the condenser was adjusted by varying the step damper position. Data for fault-free and fault conditions were collected using a PNTECH CONTROLS DDC-C46 and transmitted to a PC via a Modbus RTU 485 at 5-s intervals.

TABLE I. Methods of Implementing the Faults and Levels Simulated

Faults	Severity levels of faults				Methods of simulated faults
	Fault free	Level 1	Level 2	Level 3	
Refrigerant leakage	0%	−10%	−20%	−30%	% reduction in the total charge
Condenser fouling	0%	−10%	−20%	−30%	The surface of the condenser coil was blocked with a volume control damper

To enhance model accuracy, data quality was prioritized. Data preprocessing involved applying a moving window filter based on three standard deviations from the mean and 10 test points to remove ambiguous data. The cleaned data were then divided into training, validation, and testing sets for analysis.

C. Data Preprocessing and Fault Diagnosis Performance Evaluation Indices

Data quality is crucial for improving model accuracy in machine learning; however, clean, well-organized data are rarely available from the outset, because they often include transient and steady-state conditions. Therefore, data preprocessing is essential to convert raw data into usable format. Common preprocessing methods include data standardization and normalization. This study employs z-score normalization, a standardization technique that scales data to have a mean of 0 and a standard deviation of 1, ensuring that features share a common scale without altering the range of values. The z-score normalization is as follows:

$$z = \frac{(X-\mu)}{\sigma} \tag{9}$$

where:
Z is the normalized value. X is the input data point. μ is the mean of the dataset. σ is the standard deviation of the dataset.

D. Evaluation Metrics for Classification Models

Evaluation metrics are essential for assessing the performance and effectiveness of machine learning models, particularly in the prediction phase of this study. The key metrics widely used for classification evaluation included accuracy, confusion matrix, precision, recall, and the F1-Score, which collectively measure the model's predictive capability [17–19]. Particularly, accuracy reflects the overall proportion of correctly classified samples. Precision indicates how many of the predicted positive cases are actually positive, emphasizing the model's reliability when predicting faults. Recall (sensitivity) measures the ability of the model to detect actual fault cases, focusing on minimizing missed detections. The F1-score combines precision and recall into a single metric, offering a balanced evaluation when both false positives and false negatives are important. The detailed formulas of these evaluation metrics are presented in Table II.

According to Vujović's report [17], this study utilized the confusion matrix to evaluate model performance in fault detection and diagnosis (Table III) for the binary classification setup. Here, faults represent the positive class, and fault-free conditions represent the negative class. Each metric was calculated by comparing the model's predictions with the actual values in the training and testing datasets.

TABLE II. METRICS FORMULA

Metric	Formula
Accuracy	$(TP+TN)/(TP+FP+TN+FN)$
Precision	$TP/(TP+FP)$
Recall (sensitivity)	$TP/(TP+FN)$
F1_Score	$\frac{2 \times Precision \times Recall}{Precision + Recall}$

TABLE III. CONFUSION MATRIX

Actual label	Predicted label	
	Fault	**Fault free**
Fault	True positive (TP)	False negative (FN)
Fault free	False positive (FP)	True negative (TN)

In Table III, TP represents the count of correctly predicted positive samples, FN represents the count of positive samples incorrectly predicted as negative, TN represents the count of correctly predicted negative samples, and FP represents the count of negative samples incorrectly predicted as positive. In a multi-class classification problem such as fault diagnosis, the positive category refers to the class currently under consideration, and the negative category includes all other combined classes.

IV. RESULTS AND DISCUSSION

We trained two machine learning models using 745 data samples for validation and 568 data samples for testing. The test set included normal operating conditions and two fault types: condenser fouling and refrigerant leakage. The data covered seven distinct classes, labeled as "0" for fault-free, "1", "2", and "3" for increasing levels of condenser fouling severity, and "−1", "−2", and "−3" for increasing levels of refrigerant leakage severity. The performance results of both models are presented in Table IV. To evaluate the robustness and generalization ability of the classification models, a 10-fold cross-validation strategy was employed during model training. In this approach, the training dataset was randomly partitioned into ten equal subsets. For each iteration, one subset was retained for validation while the remaining nine were used for training.

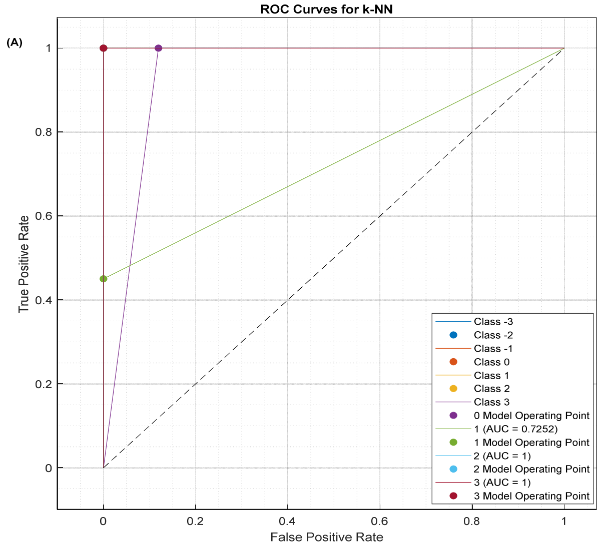
TABLE IV. METRICS RESULTS OF THE TWO MODELS

Metrics	Accuracy		Precision		Recall		F1-Score	
Model	**Validation**	**Test**	**Validation**	**Test**	**Validation**	**Test**	**Validation**	**Test**
kNN	91.68%	89.26%	92.09%	92.15%	97.98%	92.62%	93.46%	89.61%
RF	91.68%	89.26%	92.09%	92.15%	90.92%	93.59%	88.03%	90.46%

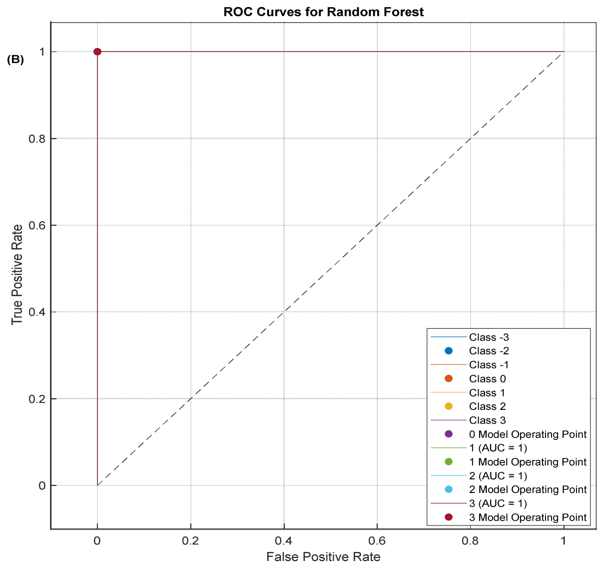
The results presented in Table IV provide a comparative performance analysis between the kNN and RF models across the validation and test sets, using accuracy, precision, recall, and the F1-Score. Both models exhibited identical validation accuracy of 91.68%, indicating similar learning performance during the training phase. However, on the test set, both models showed a slight drop in accuracy to 89.26%, suggesting a comparable ability to generalize unseen data. Precision remained consistent across both models, with identical values of 92.09% during validation and 92.15% during testing, highlighting similar behavior in identifying TPs. Differences emerge between recall and F1-Scores. The kNN model maintained a higher recall (97.98% in validation and 92.62% in testing), suggesting that it performed better in capturing all relevant instances (TPs), compared to the RF model, which showed a lower recall (90.92% in validation and 93.59% in testing). The kNN's F1-Score (93.46% in validation and 89.61% in testing) slightly declined on the test set but remained close to the RF model's F1-Score (88.03% in validation and 90.46% in testing). The key takeaway from this table is that both models perform comparably; however, the kNN model has a slight advantage in recall, whereas the RF model demonstrates more stability in its performance across different metrics, particularly in precision.

In terms of overall performance, the kNN and RF models exhibited competitive results across accuracy, precision, recall, and the F1-Score. However, each model had its strengths: the kNN model demonstrated higher recall, indicating better performance in identifying TPs, particularly during validation. The RF model demonstrated more stability across different metrics, with consistent precision and F1-Scores between the validation and test sets, which may suggest better generalizability and robustness. While the kNN model excels slightly in recall, the RF model's consistent performance across multiple metrics, especially its stable precision, suggests that it may be the better overall performer, especially in scenarios where precision and stability across different conditions are crucial. Therefore, RF slightly outperformed kNN in terms of its balanced and reliable performance.

To validate the robustness of the proposed models, we conducted a comparative performance analysis with findings reported in existing literature. Lei *et al.* [11] achieved a high F1-Score of 95.73% in refrigerant leak diagnosis using an Extremely Randomized Trees (EXT) model, which required advanced feature selection and tuning across 25 different machine learning algorithms. In contrast, our models, namely k-Nearest Neighbors (kNN) and Random Forest (RF), reached commendable F1-Scores of 89.61% and 90.46%, respectively, on the test set. Despite slightly lower F1-Scores, our approach benefits from reduced model complexity, fewer preprocessing steps, and excellent generalization across seven distinct fault and normal states. These results highlight that our models offer a more practical and efficient alternative for implementation in automotive air-conditioning fault diagnosis systems, especially where resource constraints or simplicity are prioritized. Furthermore, to assess and compare the classification models' ability to distinguish between fault conditions in vehicle A/C systems, the ROC curve was employed. This curve visually represents the trade-off between the TP and FP rates across various decision thresholds, providing a comprehensive view of each model's performance in classifying different fault conditions. By comparing the ROC curves for each model, this analysis highlights the strengths and weaknesses of each classifier, particularly their ability to handle multiple classes and generalize to unseen fault data. Fig. 3 illustrates the ROC curves of the two classification models.



(a)



(b)

Fig. 3. Evaluation of classifier performance using ROC curves (a) kNN model; (b) RF model.

Fig. 3(a) and (b) display the ROC curves for the k-NN and RF classifiers, which provide insights into the performance of these models across multiple classes. In Fig. 3(a), the k-NN classifier exhibited varying performance across different classes. For classes −3 and −2, the ROC curves indicate poor performance with high FP rates and low TP rates, suggesting that the proposed k-NN struggled to distinguish between these classes, possibly due to class overlap or imbalanced data. In contrast, classes 0, 1, and 3 demonstrate better separation, with higher TP rates and lower FP rates, indicates that k-NN performs well when the classes are more distinct. In contrast, Fig. 3(b), which shows the ROC curves for the RF classifier, demonstrates the generally superior performance. RF handles class separation better across all classes, with stronger curves that show higher TP rates and lower FP rates than k-NN. This was particularly evident for classes −2, −1, and 0, where the RF classifier consistently outperformed the k-NN. Class 3, which also performed well in the k-NN model, showed a similar high performance in the RF model, but with even sharper separation, which indicates that the RF is more robust and better at handling class overlap or noise in the data. When comparing the two classifiers, RF consistently provided more reliable and accurate results across most classes, as reflected by its stronger ROC curves. The higher area under the curve for RF further supports its superior classification ability. In contrast, the proposed k-NN, while effective in some cases, struggles with more complex class distributions and overlapping classes. Therefore, for tasks involving complex decision boundaries or requiring better generalization across multiple classes, RF model is preferred, while the k-NN may still be suitable for simpler problems with fewer class overlaps.

To further analyze the results, this study presented the results based on the confusion matrix for two cases, i.e., validation and testing. Fig. 4 illustrates the confusion matrices for the kNN and RF model in the validation case. At a glance, the classification model performs better when the numbers on the main diagonal are higher.

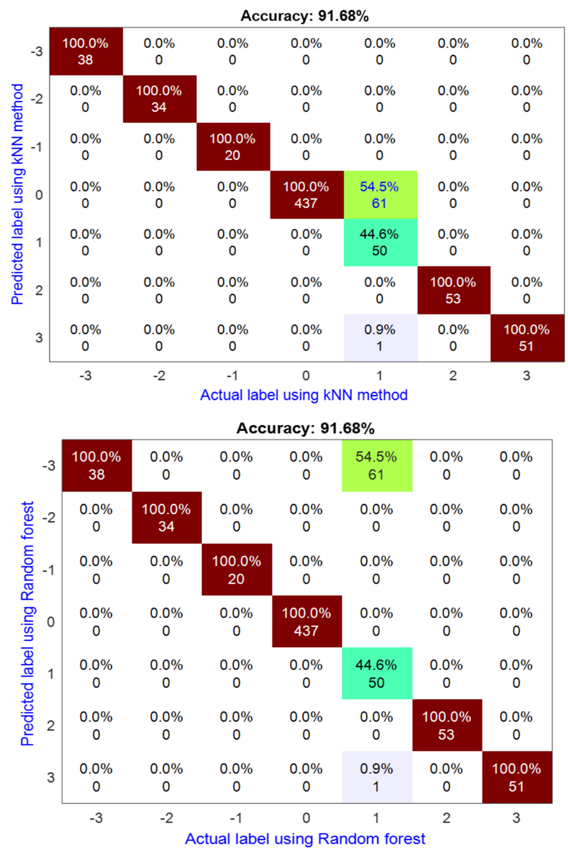


Fig. 4. Confusion matrix of the two models in validation case.

Fig. 4 presents each with an accuracy of 91.68%, for diagnosing various system conditions labeled as “0” for the fault-free state, “1”, “2”, and “3” for increasing levels of condenser fouling severity, and “−1”, “−2”, and “−3” for increasing levels of refrigerant leakage severity. For both models, the fault-free condition (label 0) and severe fault conditions (labels −3, −2, −1, and 3) were accurately identified with 100% precision, demonstrating strong model performance in distinguishing clear-cut cases. However, both models encounted difficulties with intermediate levels of condenser fouling, particularly between labels 1 and 2. The kNN model misclassified 44.6% of the label 1 instances as label 2 and 54.5% of the label 2 instances as label 1, indicating a challenge in distinguishing these similar states. The RF model exhibits an identical pattern, misclassifying the same percentage between these two labels, suggesting that the feature space overlap between mild and moderate condenser fouling levels complicates classification. Overall, both models performed well in diagnosing distinct conditions but struggled to differentiate between close severity levels, particularly in condenser fouling, highlighting an area for potential feature refinement or model enhancement.

From the perspective of fault diagnosis, the figure indicates that mild to moderate levels of condenser fouling (labels 1 and 2) are the most challenging to diagnose accurately for the kNN and RF models. The models exhibit substantial misclassification between these two levels, with 44.6% of label 1 instances incorrectly classified as label 2 and 54.5% of label 2 instances misclassified as label 1. This suggests that features representing these intermediate stages of fouling severity may overlap in the feature space, which makes it difficult for the models to distinguish between subtle variations. This issue differs from the performance of the models in more specific scenarios, such as when there are no faults (label 0) and when there are significant issues with condenser fouling (label 3) and refrigerant leakage (labels −1, −2, −3), where both models consistently perform well. It is challenging to identify these intermediate fouling conditions, indicating the need for improved feature engineering to more accurately capture the nuances of the varying severity levels.

Fig. 5, which presents the confusion matrices for the kNN and RF models on the test dataset, shows that both models achieved an overall accuracy of 89.26%. For both models, the fault-free condition (label 0) and the severe levels of refrigerant leakage and condenser fouling (labels −3, −2, −1, and 3) were accurately classified with 100% accuracy, demonstrating the models’ capability to distinguish clear fault boundaries. However, both models struggled with intermediate levels of condenser fouling severity, particularly between labels 1 and 2, similar to the validation results.

Both the kNN and RF models consistently misclassified approximately 55% of the label 1 instances as label 2 and 45% of the label 2 instances as label 1. This consistent misclassification suggests overlapping features between these fouling severity levels, thereby complicating the accurate diagnosis, even for new, unseen data. Such errors point to limitations in the models’ ability to distinguish subtle differences between mild and moderate fouling conditions, which could be due to insufficient feature differentiation for these specific fault levels.

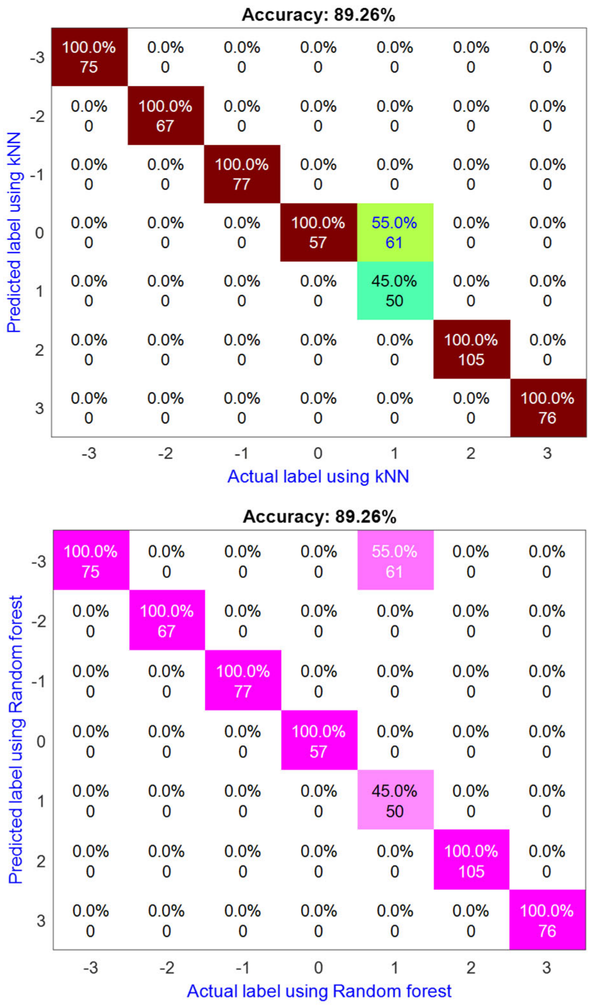


Fig. 5. Confusion matrix of two models in test case.

Overall, although both models perform well in identifying distinct fault states and severe conditions, they exhibit significant challenges in diagnosing intermediate fault severities, particularly in the early stages of condenser fouling. This highlights a potential area for model refinement, such as enhanced feature engineering or alternative modeling approaches, to improve the diagnostic accuracy of closely related fault conditions.

In summarizing both Figs. 4 and 5, the kNN and RF models exhibit similar strengths and limitations across the validation and test datasets. Both models accurately identified the fault-free condition (label 0) and severe fault levels for refrigerant leakage (labels −3, −2, −1) and condenser fouling (label 3), achieving 100% classification accuracy for these distinct states. However, both models face significant challenges in distinguishing between intermediate levels of condenser fouling severity (labels 1 and 2). In both the validation and test cases, kNN and RF showed considerable misclassification between these two levels, with approximately 55% of label 1 misclassified as label 2 and 45% of label 2 misclassified as label 1. This consistent pattern of misclassification across both datasets suggests that the feature space overlap for these mild to moderate fouling levels hinders the models' diagnostic accuracy, despite their overall high performance in distinguishing more pronounced fault conditions. Improving feature representation or exploring alternative modeling techniques could enhance the models' ability to diagnose closely related intermediate faults.

Overall, the kNN and RF models demonstrated comparable performance, with each achieving high accuracy for clearly defined conditions, such as fault-free (label 0) and severe fault levels (labels −3, −2, −1 for refrigerant leakage and 3 for condenser fouling). However, when evaluating the two models on subtle diagnostic capabilities, specifically the intermediate levels of condenser fouling severity (labels 1 and 2), both exhibited similar misclassification rates, suggesting that neither model outperforms the other in this area.

Despite this, the RF model tends to have a slight edge in overall performance owing to its consistency and robustness across various metrics in other analyses (e.g., precision and F1-Score stability). RF often generalizes better to different datasets and may be more adaptable to complex, nonlinear patterns. Consequently, although both models proved limitations in distinguishing between close severity levels, RF's robustness across metrics generally makes it the preferable choice, particularly if consistent performance across diverse conditions is prioritized.

V. Conclusion

This study evaluated and compared the performance of two machine learning models, kNN and RF, for fault diagnosis in vehicle A/C systems using a dataset of 745 validation samples and 568 test samples. The data cover seven distinct fault conditions, including normal operating conditions and varying severities of condenser fouling and refrigerant leakage. Both models demonstrated strong performance, achieving an identical validation accuracy of 91.68%, with a slight decrease to 89.26% for the test set. Precision remained consistent across both models, whereas differences emerged in recall and F1-Scores.

The kNN model outperformed the RF model in recall, particularly in capturing TPs, while the RF model proved superior stability and consistency across all performance metrics, including precision and F1-Score. The evaluation using ROC curves further supported the RF model's robustness, with stronger classification ability across all fault conditions compared to kNN. The confusion matrix analysis highlighted both models' strengths in diagnosing fault-free and severe conditions but revealed significant misclassification between intermediate levels of condenser fouling, suggesting that improved feature differentiation is required for better diagnosis of these subtle fault variations.

Despite both models performing well in identifying clear fault boundaries, their inability to effectively distinguish between mild and moderate condenser fouling (labels 1 and 2) remains a challenge. This issue underscores the need for further refinement in feature engineering or the exploration of alternative modeling approaches to improve the classification accuracy of overlapping fault conditions.

Overall, while kNN and RF models exhibit similar strengths in fault diagnosis, the RF model's superior generalization and stability across different performance metrics make it a more reliable choice, particularly in scenarios where consistent performance and robustness are crucial. This study highlights the importance of model choice depending on the specific requirements of fault diagnosis systems, with RF emerging as the preferred option for handling complex fault conditions with better adaptability to unseen data.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTION STATEMENT

Authors Dinh Anh Tuan Tran, Thi Khanh Phuong Ho and Van Tuan Nguyen proposed the research problem developed the theory, and performed the computations. All authors have discussed the results and contributed to the final manuscript. All authors had approved the final version.

REFERENCES

[1] Ü. Ağbulut, "Forecasting of transportation-related energy demand and CO_2 emissions in Turkey with different machine learning algorithms," *Sustain Prod Consum*, vol. 29, pp. 141–157, 2022. https://doi.org/10.1016/j.spc.2021.10.001

[2] K. Shete, "Influence of automotive air conditioning load on fuel economy of IC engine vehicles," *Int. J. Sci. Eng. Res.*, vol. 6, no. 8, pp. 1367–1372, 2015.

[3] Y. Chen, D. A. T. Tran, H. L. Ao, and H. N. T. Cam, "An enhanced chiller FDD strategy based on the combination of the LSSVR-DE model and EWMA control charts," *International Journal of Refrigeration*, vol. 72, pp. 81–96, 2016. https://doi.org/10.1016/j.ijrefrig.2016.07.024

[4] F. Xiao, C. Zheng, and S. Wang, "A fault detection and diagnosis strategy with enhanced sensitivity for centrifugal chillers," *Appl. Therm. Eng.*, vol. 31, no. 17–18, pp. 3963–3970, Dec. 2011. https://doi.org/10.1016/J.APPLTHERMALENG.2011.07.047

[5] J. Cui and S. Wang, "A model-based online fault detection and diagnosis strategy for centrifugal chiller systems," *International Journal of Thermal Sciences*, vol. 44, no. 10, pp. 986–999, 2005. http://dx.doi.org/10.1016/j.ijthermalsci.2005.03.004

[6] B. R. Vasconcellos, M. Rudek, and M. de Souza, "A machine learning method for vehicle classification by inductive waveform analysis," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 13928–13932, 2020. https://doi.org/10.1016/j.ifacol.2020.12.908

[7] M. R. Bachute and J. M. Subhedar, "Autonomous driving architectures: insights of machine learning and deep learning algorithms," *Machine Learning with Applications*, vol. 6, 100164, 2021. https://doi.org/10.1016/j.mlwa.2021.100164

[8] K. Rana and P. Kaur, "Review on machine learning based algorithms used in autonomous cars," *International Journal of Emerging Technologies and Innovative Research*, vol. 5, Oct. 2018.

[9] N. Saqib, M. Yousuf, and M. Rashid, "Design and implementation issues in autonomous vehicles—A comparative review," in *Proc. 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, Dubai, United Arab Emirates, 2021. https://doi.org/10.1109/ICCAKM50778.2021.9357706

[10] D. A. T. Tran, Y. Chen, M. Q. Chau, and B. Ning, "A robust online fault detection and diagnosis strategy of centrifugal chiller systems for building energy efficiency," *Energy Build*, vol. 108, pp. 441–453, 2015. https://doi.org/10.1016/j.enbuild.2015.09.044

[11] Q. Lei, C. Zhang, J. Shi, and J. Chen, "Machine learning based refrigerant leak diagnosis for a vehicle heat pump system," *Appl. Therm. Eng.*, vol. 215, 118524, 2022. https://doi.org/10.1016/j.applthermaleng.2022.118524

[12] Y. Xia, J. Zhao, Q. Ding, and A. Jiang, "Incipient chiller fault diagnosis using an optimized least squares support vector machine with gravitational search algorithm", *Front. Energy Res.*, vol. 9, 755649, 2021. https://www.frontiersin.org/article/10.3389/fenrg.2021.755649

[13] K. Yan, W. Shen, T. Mulumba, and A. Afshari, "ARX model based fault detection and diagnosis for chillers using support vector machines," *Energy Build*, vol. 81, pp. 287–295, 2014. https://doi.org/10.1016/j.enbuild.2014.05.049

[14] C. Youming, J. Changliang, and T. D. A. Tuan, "A fault detection and diagnosis method based on kriging model for chillers," *Journal of Hunan University Natural Sciences*, vol. 43, no. 9, pp. 70–80, Sep. 2016.

[15] D. Dehestani, S. Su, H. Nguyen, and Y. Guo, "Robust fault tolerant application for HVAC system based on combination of online SVM and ANN black box model," in *Proc. 2013 European Control Conference (ECC)*, 2013, pp. 2976–2981. https://doi.org/10.23919/ECC.2013.6669140

[16] D. A. T. Tran, H. N. T. Cam, and Q. M. Phan, "A robust neural network algorithm for automotive air conditioning fault detection," *Int. J. Automot. Eng.*, vol. 11, no. 1, pp. 9–13, 2020. https://doi.org/10.20485/jsaeijae.11.1_9

[17] Ž. Vujović, "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021. https://doi.org/10.14569/IJACSA.2021.0120670

[18] M. G. Akpan, U. D. George, and D. N. Onwe, "Machine learning approach for classification of dalium guineense fruits," *World Journal of Applied Science & Technology*, vol. 15, no. 1, pp. 38–42, Nov. 2023. https://doi.org/10.4314/wojast.v15i1.38

[19] M. Vihinen, "How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis," *BMC Genomics*, vol. 13, suppl. 4, 2012. https://doi.org/10.1186/1471-2164-13-S4-S2

Copyright © 2025 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).