International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 07 July 2025, pp: 677-683 www.ijaem.net ISSN: 2395-5252

Machine Learning Model for Detecting Faults in Variable Refrigerant Flow (VRF) Systems

Garba Muhammad, Mohammed Ali Kawo, Shamsudeen Lawal
[1,2,3]*Department of Computer Science, Federal University Gusau, Zamfara State, Nigeria*
Corresponding Author: Mohammed Ali Kawo

Date of Submission: 15-07-2025 Date of Acceptance: 25-07-2025

ABSTRACT: This study focuses on building a machine learning model to forecast maintenance requirements for Variable Refrigerant Flow (VRF) systems by examining their operational data. A detailed dataset, obtained from Kaggle.com, includes essential parameters such as temperature, pressure, compressor behaviour and energy usage along with maintenance and failure logs. The data is thoroughly preprocessed to enhance its accuracy and dependability, followed by exploratory analysis to identify key patterns and correlations. Predictive models like Random Forest (RF), Support Vector Machine (SVM) and K-Nearest Neighbour (KNN) are evaluated to determine which offers better classification performance. These models are then trained and validated using Accuracy score, F1-score, Precision and Recall respectively.

KEYWORDS: Random Forest;Support Vector Machine;Variable Refrigerant Flow; machine learning.

I. INTRODUCTION

The need for efficient heating, ventilation, and air conditioning (HVAC) solutions has increased dramatically recently, owing to rising energy costs and increased awareness of environmental sustainability. Variable Refrigerant Flow (VRF) systems have become popular due to their energy efficiency, versatility, and ability to accurately control temperature across multiple zones.(Yau et al. (2024), large capacity ductless multi-split air conditioning system versions are called VRF systems, connected to a range of interior units with a fan, evaporator, expansion device, and controls, they comprise an outdoor unit with one or more variable speed inverter/stepped compressors, condenser, accumulator, receiver, and components.

However, VRF systems, like any other complicated system, require regular maintenance to ensure durability and peak performance. As it has been done historically, routine maintenance frequently depends on planned inspections and reactive measures to address system faults. This strategy may result in inefficient maintenance checks and unscheduled downtime when problems happen without warning. Using operational data and failure histories, predictive maintenance models provide a proactive approach to these problems by predicting when repair is necessary before problems worsen.

An HVAC (heating, ventilation, and air conditioning) system that uses variable refrigerant flow (VRF) technology offers buildings extremely effective and adaptable climate control. Large homes, workplaces, and industrial buildings are among the settings where it works best since it provides accurate temperature management. In fact, a VRF system has a number of benefits that make it a well-liked and effective substitute for centralized HVAC systems. Zone-wise temperature management, optimal performance under part load, and reduced installation area requirements are some of the main advantages.

The inability to quickly and accurately identify faults in VRF systems, which results in significant energy waste, compromised occupant comfort, increased maintenance costs, and the potential for premature equipment failure, is one of the major issues that still need to be fixed despite the system's many benefits.

By using a comprehensive approach, predictive maintenance plans are guaranteed to be grounded in both the overall operational context and the performance of individual machines, resulting in more precise and efficient maintenance decisions (Kim et al., 2017).

II. LITERATURE REVIEW

According to Barney et al. (2021), to improved maintenance results and long-term competitive advantage can result from industrial firms prioritizing the deployment and integration of machine learning models in predictive maintenance by understanding their strategic worth.

Lee, Bagheri & Kao (2014) examined the cyber-physical system (CPS) architecture for Industry 4.0 manufacturing systems' predictive maintenance. The researchers developed a predictive maintenance framework by combining big data analytics, machine learning, and the Internet of Things in a CPS architecture. The framework was put to the test in a manufacturing simulation. Because the CPS architecture allowed for real-time equipment data monitoring and analysis, it greatly increased the accuracy of maintenance predictions and decreased downtime. By incorporating machine learning models, the system's predictive power was improved.

According to (Yau et al. (2024), a review of multi-split VRF system configuration in depth was carried out, comparing its applicability and operating performance to that of standard AC systems. Research on controlling compressor frequency and electronic expansion valve (EEV) opening simultaneously was suggested. This study falls short in the development of the VRF system with other integrated systems since it focuses more on the control side.

According to Kim et al., (2021), using a thorough approach ensures that predictive maintenance plans are based on the overall operation and the performance of each machine, leading to better and more effective maintenance choices.

Chandola, Banerjee & Kumar (2017) aimed to offer a thorough rundown of machine learning-based anomaly detection methods for predictive maintenance. They discovered that parts that suggest possible equipment breakdowns were especially well-identified using unsupervised learning techniques, like clustering algorithms. Though they needed a lot of labelled data, supervised models like KNNs and decision trees also showed promise.

Lee, Bagheri & Kao (2014) investigated the design of cyber-physical systems (CPS) for Industry 4.0 manufacturing systems' predictive maintenance. In order to develop a predictive maintenance framework, the researchers suggested a CPS architecture that combines IoT, big data analytics, and machine learning.The CPS framework made it possible to monitor and analyse equipment data in real-time, which greatly increased the accuracy of maintenance predictions and decreased downtime. The incorporation of machine learning models improved the system's capacity for prediction.

In this research, the authors used different parameters to evaluate performance of classifiers such as Random Forest (RF), Support Vector Machine and K-Nearest Neighbor (KNN) in fault detection of Variable Refrigerant Flow (VRF) systems.

III. PROPOSED METHODOLOGY

In this segment some important phases are well described.

A. Collection of Dataset, weretrievedan online data from kaggle.com a machine learning repository. Data is cleaned, preprocessed, and prepared for model training

B. Model Development and Training, an appropriate machine learning models such as Support Vector (SVM), K-Nearest Neighbour (KNN) and Random Forest (RF) are selected and trained on the prepared dataset.

i. Support Vector Machine

A supervised learning technique for both linear and nonlinear binary classification is Support Vector Machines (SVM). (Arya et al., 2019), Some of the better locations that are close to the surface are chosen by the classifier. The chosen spots, which show both positive and negative samples, are located close to the surface. Using a hyperplane, the vectors document is separated into two classes. Finding the optimal decision boundary to split data points into two distinct classes is the goal of support vector machines.

ii. K-Nearest neighbor

The K-nearest neighbor algorithm is virtually used in regression and classification. It identifies unknown samples from the class, looks for the k-closest cases, and builds a prediction based on the majority, where the majority belongs to the closest neighbours. The algorithm's functionality is very straightforward and successful in correctly classifying textual documents.

iii. Random Forest

For each patient observation, the Random Forest Classifier uses a number of decision trees. The algorithm begins by using the training set's

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 07 July 2025, pp: 677-683 www.ijaem.net ISSN: 2395-5252

random observations. The algorithm then builds a decision tree for each of these observations, yielding appropriate findings. Each variable is predicted, and the ultimate prediction result for that observation is decided by voting (Obaidat et al., 2022).

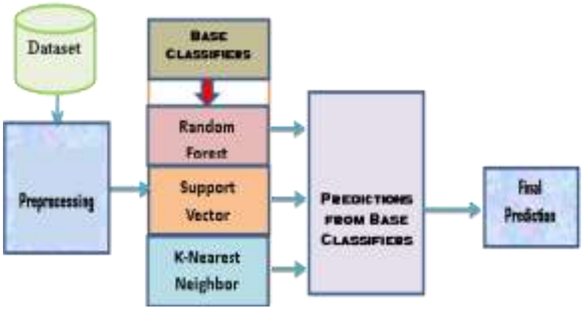

Fig 1: Proposed system architecture

C. Detection and Predict Analysis: The trained model analyses real-time data and detects the likelihood of future failures or the need for maintenance.In order to improve maintenance schedules, increase efficiency, and reduce disruptions, this research makes use of machine learning approaches to improve the defect identification of VRF systems.

D. Model validation and testing, machine learning model performance are normally evaluated through the model's performance. The effectiveness in meeting its goals, the impact of VRF system performance and maintenance costs and to evaluate user satisfaction in respect to the model improvements.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental Setting

We performed the entire experiments on a HP laptop system AMD Ryzen 3, 3250U with Radeon 2.60Ghz graphics, 8 GB RAM, on Windows 10 Pro, X64 bits-based processor version 22H2. The open-source Python distributor Jupyter Notebook (Version 6.5.4 IDE) from Anaconda is the language used to conduct the experiments in this study. Panda and NumPy are used to handle the datasets and dataframe, while scikit-learn, matplotlib, and seaborn are used for training, testing, mathematical computation, and visualization of the full results.

B. Dataset

The following characteristics and labels were applied to the industrial fault detection dataset that was obtained from kaggle.com. This dataset is intended for use with IoT-based smart sensors for automatic fault detection in VRF applications. It incorporates industrial sensor data gathered from multiple sources, including defects like: compressor fault, power fluctuation fault, refrigerant leakage and normal operation. These faults are categorized into numbers from zero to three respectively.

C. Evaluation Parameters

To assess the suggested method's performance, we employed a confusion matrix. Once the classifiers have classified the data sets, the fundamental attitudes are generated. Several classification metrics, including accuracy, F1_score, precision, and recall, are used to compare the model's performance against baseline methods.

D. Performance Interpretation

The experimental results from Tables 1 through figure 3 summarize the classification methods that were used on Random Forest (RF),Support Vector Machine (SVM) and K-Nearest Neighbor (KNN).Table 1 below depicts the classification results of the three models on industrial fault detection dataset in respect of accuracy score, precision, recall and F1-score.
Based on the table 1 below, Support Vector Machine (SVM) outperformed all other classifiers with a higher accuracy score of 70% with a precision score of another 70%, an ultimate recall score of 100% and 82% as F1-score respectively. The Random Forest (RF) and K-Nearest Neighbor (KNN) classifiers came next with an accuracy score of 69%.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 07 July 2025, pp: 677-683 www.ijaem.net ISSN: 2395-5252

Table 1: Classification result of three models

Model	Accuracy	Precision	Recall	F1-score
Random Forest	69%	70%	99%	82%
S-V-Machine	70%	70%	100%	82%
K-N-Neighbor	69%	70%	99%	82%

From all indications, Support Vector Machine (SVM) is the suitable model for the fault detection task.

i. Confusion Matrix Analysis

The confusion matrix for the Support Vector Machine model is shown below:

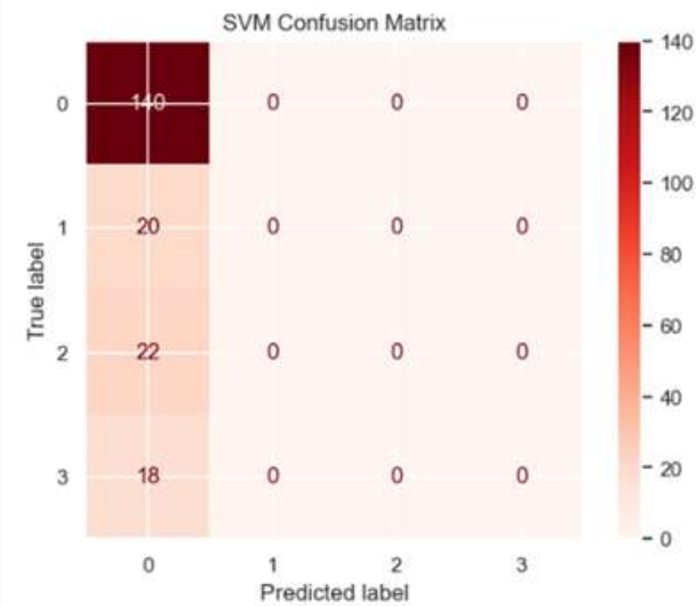

Fig 2: Confusion Matrix of Support Vector Machine Model

The confusion matrix above displays the fault types ranging from 0, 1, 2, and 3. Where zero indicate none fault, one designates a compressor fault, two a refrigerant leakage fault and three indicates power fault. The model shows strong classification ability with minimal misclassifications between similar fault types.

ii. Bar Chart Analysis for the Models

The Bar chart for each classifier indicate tremendous performance for all fault classes with a high sensitivity and specificity. The Support vector machine (SVM) has shown a robust enhancement in recall with a 100% score even though the accuracy score supersedes all other classifiers in term of detecting faults. Figure 3 below display bar chart for the three classifiers on industrial fault detection of VRF dataset.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 07 July 2025, pp: 677-683 www.ijaem.net ISSN: 2395-5252

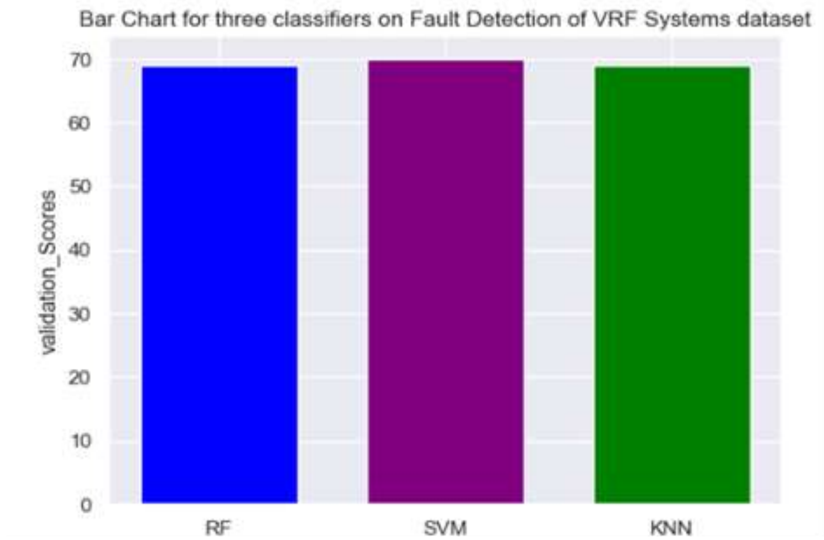

Fig 3: Bar Chart for Fault Detection on VRF Dataset

E. Results Discussion

The results indicate that machine learning, especially Support Vector Machine, can effectively detect and classify faults in VRF systems with high reliability. The minimal false positive rate is critical in preventing unnecessary maintenance, while high recall ensures that genuine faults are not missed.

The deployment of a machine learning-based fault detection system for VRF HVAC systems shows how contemporary data-driven methodologies may greatly improve system reliability and fault diagnosis. The system uses sensor data and intelligent categorization models to diagnose HVAC problems faster and with less skill, especially in complicated systems like VRFs that have several interior units and dynamic refrigerant flow control.

Among the most noteworthy results are the following:

• Reduced problem discovery Time: Automated warnings and real-time monitoring reduced the problem discovery time from hours of manual diagnostic to minutes.

• High Model Accuracy: Random Forest (RF) and K-Nearest Neighbor (KNN) models each had an average accuracy score of 69%, while Support Vector Machine (SVM) did better with an accuracy score of 70%. This demonstrates their suitability for multivariate fault classification tasks in VRF systems.

• Scalability: The system's modular architecture enables it to be integrated with current Building Management Systems (BMS) and tailored to various VRF configurations.

V. CONCLUSION

This research objective is to utilize machine learning technique to successfully demonstrates the feasibility and effectiveness of using Randon Forest (RF), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) for fault detection in Variable Refrigerant Flow HVAC systems. Through structured data collection, preprocessing, feature engineering, and model evaluation, a reliable diagnostic tool was developed and validated.

The technology helps save energy by preventing performance degradation from undiagnosed problems, in addition to increasing operational efficiency and decreasing downtime.

VI. FUTURE WORK

There are numerous ways to build on this research and improve the system even more:

- More Diverse and Larger Datasets: Model generalization will be enhanced by working with facilities managers or HVAC manufacturers to obtain a larger dataset with a wider range of fault scenarios.
- Hybrid Models: Especially for unclear failure scenarios, combining machine learning with rule-based or physics-based models can improve dependability
- Online Learning: Using models that can adjust to fresh data in real time, like reinforcement learning will help control concept drift and guarantee correctness over time.
- Anomaly Detection: To find new information, unsupervised learning techniques could be used.

REFERENCES

[1]. Arya, P., Bhagat, A., & Nair, R. (2019). Improved Performance of Machine Learning Algorithms via Ensemble Learning Methods of Sentiment Analysis. 10(2), 110–116.

[2]. Barney, J. B., Ketchen, D. J., & Wright, M. (2021). Resource-Based Theory and the Value Creation framework. Journal of Management, 47(7), 1936–1955. https://doi.org/10.1177/01492063211021655

[3]. Boyes H., Hallaq B., Cunningham J., & Watson T. (2018). The industrial internet of things (IIoT): An analysis framework, Compute. Ind., 101, 1–12. https://doi.org/10.1016/j.compind.2018.04.015

[4]. Chandola, V., Banerjee, A., & Kumar, V. (2017). Active learning. In Encyclopedia of Machine Learning and Data Mining (pp. 42–56). https://doi.org/10.1007/978-1-4899-7687-1_912.Cho H. and Lee B. (2011). A fault detection and diagnosis method for VRF systems using artificial neural networks and fuzzy logic, International Journal of Refrigeration, vol. 34, no. 1, pp. 202-211.

[5]. Gao Z., Cecati C., & Ding S. X. (2015). A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches, IEEE Trans. Ind. Electron., 62, 3757–3767. https://doi.org/10.1109/TIE.2015.2417501

[6]. Isermann R. (2005) Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance, Springer Science & Business Media. Jordan M. I., & Mitchell T. M. (2015). Machine learning: Trends, perspectives, and prospects, Science, 349, 255–260. https://doi.org/10.1126/science.aaa8415

[7]. Kim, D., Cox, S. J., Cho, H., & Im, P. (2017). Evaluation of energy savings potential of variable refrigerant flow (VRF) from variable air volume (VAV) in the U.S. climate locations. Energy Reports, 3, 85–93. https://doi.org/10.1016/j.egyr.2017.05.002.

[8]. Kim K. & Lee W. (2007). Fault detection and diagnosis of air-handling units using rule-based and neural network approaches, Applied Thermal Engineering, vol. 27, no. 1, pp. 193-201.

[9]. Lee, J., Bagheri, B., & Kao, H. (2014). A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. Manufacturing Letters, 3, 18–23. https://doi.org/10.1016/j.mfglet.2014.12.001.

[10]. Li S., Wang Y., & Zhang X. (2018). A novel fault detection and diagnosis method for VRF systems based on dynamic time warping and Bayesian network, Applied Thermal Engineering, vol. 128, pp. 290-300.

[11]. Lu L., Sangaiah S., & Xie M. (2019). A survey on deep learning approaches for fault diagnosis, Engineering Applications of Artificial Intelligence, vol. 86, pp. 115-126.

[12]. Mohammadi M., движения A., & движения, S. (2019). A data-driven fault detection and diagnosis framework for HVAC systems using machine learning techniques," Building and Environment, vol. 162, p. 106277.

[13]. Muller A., Marquez A. C., & Iung B. (2008). On the concept of e-maintenance: Review and current research, Reliable. Eng. Syst. Saf., 93, 1165–1187. https://doi.org/10.1016/j.ress.2007.08.006

[14]. Obaidat, M. A., Alexandrou, A., & Sanacore, S. (2022). Machine Learning Stacking Ensemble Model for Predicting Heart Attacks Machine Learning Stacking Ensemble Model for Predicting Heart Attacks. August.

[15]. Sahli H., (2020). An introduction to machine learning, in TORUS 1-Toward an Open Resource Using Services: Cloud Computing for Environmental Data, 61–74. https://doi.org/10.1002/9781119720492.

[16]. Shin J., & Jun H. (2015). On condition-based maintenance policy, J. Comput. Des. Eng., 2, 119–127. https://doi.org/10.1016/j.jcde.2014.12.006

[17]. Stenström C., Norrbin P., Parida A., & Kumar U. (2016). Preventive and corrective maintenance-cost comparison and cost-benefit analysis, Struct. Infrastructure. Eng., 12, 603–617. https://doi.org/10.1080/15732479.2015.1032983

[18]. Susto G. A., Schirru A., Pampuri S., & McLoone S. (2014). A. Beghi, Machine learning for predictive maintenance: A

multiple classifier approach, IEEE Trans. Ind. Inf., 11, 812–820. https://doi.org/10.1109/TII.2014.2349359.

[19]. Wang Y., Li S., & Wang C. (2015). A review on fault detection and diagnosis methods for building HVAC systems, Renewable and Sustainable Energy Reviews, vol. 41, pp. 625-641.

[20]. Yau, Y. H., Rajput, U. A., & Badarudin, A. (2024). A comprehensive review of variable refrigerant flow (VRF) and ventilation designs for thermal comfort in commercial buildings. Journal of Thermal Analysis and Calorimetry, 149(5), 1935–1961. https://doi.org/10.1007/s10973-023-12837-3.

[21]. Zhao J., Li Z., & Jiang Y. (2016). Fault detection and diagnosis of variable refrigerant flow (VRF) systems based on wavelet packet transform and support vector machine, Energy and Buildings, vol. 111, pp. 299-307.