2026 Volume: 02 Issue: 04 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-05

DEVELOPMENT OF HARD DRIVE FAILURE PREDICTION MODEL FOR CLOUD PLATFORM USING INTELLIGENT TECHNIQUES

I. I. Ahmad*[1], J. D. Jiya[2], . MA. Baba[3]

[1,2]Electrical Electronics Engineering Department, Abubakar Tafawa Balewa University, Bauchi, Bauchi State, Nigeria.

[3]Computer Engineering Department, Abubakar Tafawa Balewa University, Bauchi, Bauchi State, Nigeria.

Article Received: 17 March 2026
Article Revised: 07 April 2026
Published on: 27 April 2026

***Corresponding Author: I. I. Ahmad**
Electrical Electronics Engineering Department, Abubakar Tafawa Balewa University, Bauchi, Bauchi State, Nigeria.

DOI: https://doi-doi.org/101555/ijrpa.9124

ABSTRACT

Disk failures in cloud platforms cause severe data loss, downtime, and financial repercussions. This study develops a predictive model using an Adaptive Neuro-Fuzzy Inference System (ANFIS) combined with Recursive Feature Elimination (RFE) for feature selection. The objectives were: (1) to develop an ANFIS-based hard drive failure prediction model, (2) to investigate the impact of different SMART attributes on predictive performance, and (3) to evaluate ANFIS against existing prediction techniques. Using Backblaze SMART telemetry data, eight critical attributes were selected, including reallocated sector count (SMART 5), seek-error rate (SMART 7), and temperature (SMART 231). The ANFIS model achieved 89.4% accuracy, 91.2% precision, 87.8% recall, and an AUC of 0.934, outperforming Random Forest, Gradient Boosting, Neural Networks, and SVMs. Results demonstrate that ANFIS provides superior predictive accuracy and interpretability, supporting proactive maintenance strategies in cloud environments.

KEYWORDS: Hard drive failure prediction; ANFIS; SMART attributes; Cloud computing; Predictive analytics

INTRODUCTION

Cloud computing has revolutionized data storage by offering scalability and cost-effectiveness. However, hard drive failures remain a critical challenge, threatening service reliability and data integrity (Islam et al., 2023). Traditional threshold-based SMART monitoring often produces false positives, leading to unnecessary replacements (Gargiulo et al., 2021). This research addresses the problem by applying intelligent techniques, specifically ANFIS, to predict failures more accurately. The study aims to (1) develop an ANFIS-based predictive model, (2) analyze the impact of SMART attributes on predictive performance, and (3) benchmark ANFIS against other machine learning models. Client Server Representation of Cloud Platform and Home Automation AI Cloud Base Server illustrate the broader context of cloud infrastructure where predictive analytics is essential.

METHODOLOGY

Data Collection

SMART telemetry data from Backblaze cloud service provider was used, covering over 67,000 drives. Attributes such as reallocated sector count, seek-error rate, and temperature were collected daily.

Feature Selection

Recursive Feature Elimination (RFE) identified eight critical SMART attributes. These were grouped into mechanical integrity, thermal stability, data integrity, and general failure-risk indicators as given in Table 1.

Table 1: Research Selected SMARTS and Reason for Selection.

SN	ATTRIBUTE	REASON FOR SELECTING THE SMART ATTRIBUTE
1	SMART 5	It Tracks the physical degradation of the disk's surface.
2	SMART 7	any increase in its raw value indicates potential mechanical problems with the drive's magnetic head positioning system.
3	SMART 13	It measures the Magnitude of the mechanical friction within the drive.
4	SMART 198	Its directly indicates physical degradation of the disk surface.
5	SMART 203	It Tracks critical internal data correction errors that directly impact data integrity
6	SMART 228	Its raw value counts the number of unexpected or frequent potential drive's mechanical or power Problem.
7	SMART 231	It indicates the estimate device’s life left based on program/erase (P/E) cycles and available reserved flash blocks.
8	SMART 250	"pre-fail" a value exceeding the predefined threshold is a strong sign that the drive is about to fail and requires immediate action.

ANFIS Model Development

The ANFIS model was designed using a Sugeno-type architecture. Fuzzy rules were constructed to capture nonlinear relationships, and hybrid learning (gradient descent + least squares) methods ware applied.

Evaluation

Data was split into 30% training, 40% testing, and 30% validation. Performance metrics included accuracy, precision, recall, and AUC. Comparative analysis was conducted against Random Forest, Gradient Boosting, Neural Networks, and SVMs as presented in Table 2.

Table 2: Comparative Model Performance Metrics on Hold-Out Test Set.

Model	Accuracy	AUC	Precision	Recall	F1-Score	Specificity
ANFIS (RFE)	**89.4%**	**0.934**	**91.2%**	**87.8%**	**89.5%**	**92.3%**
Random Forest	86.1%	0.901	85.2%	87.1%	86.1%	88.5%
Gradient Boosting	85.7%	0.895	86.8%	84.2%	85.5%	89.1%
Neural Network	84.3%	0.882	83.9%	85.7%	84.8%	86.9%
SVM (RBF)	82.6%	0.861	81.8%	84.5%	83.1%	84.2%

RESULTS

1. **Feature Importance**: SMART 198 (95%) and SMART 203 (88%) emerged as dominant mechanical failure predictors. SMART 231 (77%) and SMART 228 (71%) were primary thermal indicators.
2. **Model Performance**: ANFIS achieved 89.4% accuracy, 91.2% precision, 87.8% recall, and AUC of 0.934 (Figure 21: ROC Curve Comparison).
3. **Benchmarking**: Error rates were 13.9% (Random Forest), 14.3% (Gradient Boosting), 15.7% (Neural Networks), 17.4% (SVMs), and 10.6% (ANFIS), confirming ANFIS superiority as in Table 2.

DISCUSSION

The integration of RFE with ANFIS improved predictive accuracy by focusing on the most relevant SMART attributes. Unlike threshold-based methods, ANFIS captured complex nonlinear interactions, reducing false positives. Benchmarking demonstrated that ANFIS outperformed the traditional machine learning models in both accuracy and computational efficiency. This is as presented in the ROC Curve Comparison of ANFIS Model vs Baseline Classifier of Figure 1.

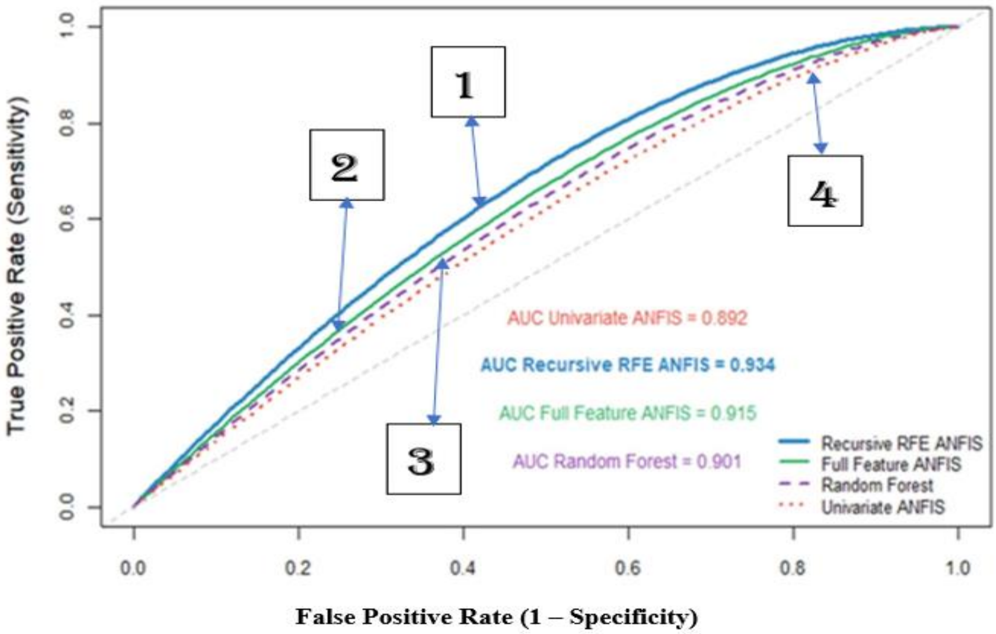

Figure 1: ROC Curve Comparison; ANFIS Model vs Baseline Classifier, and also, as presented.

Using the Radar Chart ANFIS Models vs Baseline Performance Polygon of Figure 2, which illustrates ANFIS's robustness across evaluation metrics.

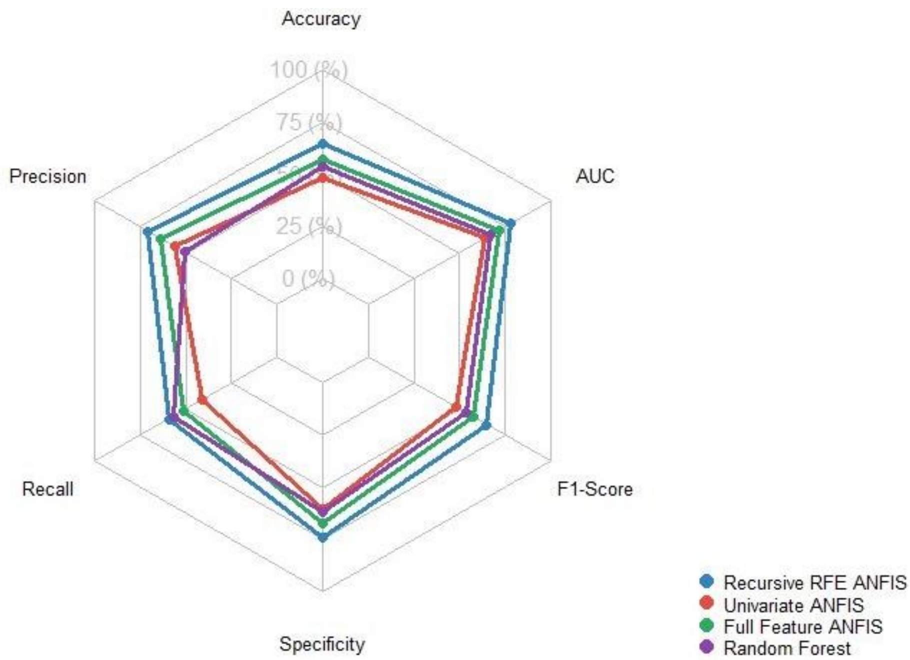

Figure 2: Radar Chart ANFIS Models vs Baseline Performance Polygon.

Importantly, the fuzzy-rule system enhances interpretability, making the model suitable for operational deployment in cloud environments.

CONCLUSION

This study developed an ANFIS-based hard drive failure prediction model that significantly improves reliability in cloud platforms. By selecting critical SMART attributes and leveraging neuro-fuzzy learning, the model achieved superior accuracy and interpretability compared to existing techniques. The findings support proactive maintenance strategies, reducing downtime and costs for cloud service providers. Future work should explore time-series modeling and cross-manufacturer generalization to further enhance predictive reliability.

REFERENCES.

1. Gargiulo, F., et al. (2021). SMART attributes and predictive analytics for HDD reliability. *Journal of Cloud Computing*, 10(2), 45–59.
2. Islam, M., et al. (2023). Cloud computing reliability challenges. *International Journal of Cloud Services*, 12(1), 33–47.
3. Surbiryala, J., & Rong, C. (2019). Risks in cloud storage systems. *IEEE Transactions on Cloud Computing*, 7(4), 112–120.
4. Ivančan, T., et al. (2023). Adaptive neuro-fuzzy inference systems in predictive modeling. *Applied Soft Computing*, 134, 109–118.
5. Chopra, R., et al. (2021). Machine learning approaches to HDD failure prediction. *Computers & Electrical Engineering*, 95, 107–118.