Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 6, December 2026, pp. 5474~5485
ISSN: 2302-9285, DOI: 10.11591/eei.v15i6.13905

Agentic AI-driven framework for transformer fault prediction, diagnosis, and autonomous rectification planning

P. Lakshmi Narayana[1], Robbi Rahim[2], Kavitha Manoharan[3], A. V. Prabu[4]
[1]Department of Electrical and Electronics Engineering, Vignan's Foundation for Science, Technology and Research, Vadlamudi, India
[2]Department of Management, Sekolah Tinggi Ilmu Manajemen Sukma, Medan, Indonesia
[3]Department of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, India
[4]Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India

Article Info

Article history:

Received Jul 7, 2026
Revised Aug 27, 2026
Accepted Sep 12, 2026

Keywords:

Agentic artificial intelligence
Autonomous rectification
Fault diagnosis
Predictive maintenance
Smart grid
Transformer fault prediction

ABSTRACT

Fault management in transformers is critical in ensuring the reliability in the power system and avoiding equipment failures. Nevertheless, the traditional monitoring methods tend to deal with fault prediction, diagnosis and corrective action independently and do not provide a combined decision making and post-action assessment. In this paper, an agentic artificial intelligence (AI) system is presented that combines transformer fault prediction, multi-class diagnosis, and autonomous rectification planning in a closed-loop system. An electrical, thermal and dissolved-gas variables reproducible synthetic dataset that includes 5,000 balanced operating records is created. Extra trees regressor is used to estimate the probability of fault and a random forest model is used to classify fault types. An expert-crafted policy based on a rule and operational and safety constraints is based on an automatic choice of corrective actions, without reinforcement learning (RL). A train-test split (80:20) and a five-fold cross-validation are used. In the held-out test set, the predictor results in MAE=0.0205, RMSE=0.0271, and R2=0.9881 and the diagnosis model results in 97.63% accuracy and macro F1=0.9762. The rule-based rectification mechanism generates an average simulated risk reduction (RR) of 72.67 with a range of 65.15-78.20 between the fault scenarios which shows integrated prediction, diagnosis, constrained action selection and feedback evaluation.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Kavitha Manoharan
Department of Electronics and Communication Engineering, Saveetha School of Engineering
Saveetha Institute of Medical and Technical Sciences, Saveetha University
Chennai, India
Email: mkseee2007@outlook.com

1. INTRODUCTION

In a power-distribution network, short circuits, thermal overloading, degradation of insulation, and non-standard loading can lead to the loss or disruption of service or economic losses caused by the distribution transformer, which is an important asset in a power-distribution network [1]. More recent research has shown the significance of multi-source data fusion to monitor condition of transformers and make intelligent fault diagnosis [2]. Dissolved-gas analysis and asset-management techniques can be used to obtain valuable information that may help in the detection of transformer degradation and the evaluation of the equipment and its criticality [3]. The deep-learning algorithms have also enhanced the detection of faults by the characteristics of insulating-oil [4]. Electronic systems fault-tolerant routing and recovery algorithms provide more comprehensive views on reliability, but have minimal application to transformer protection [5].

Predictive maintenance which is artificial intelligence (AI)-based has also been explored in fault detection and optimization of operations in smart-grids [6]. Traditional dissolved-gas analysis has its role in determining faults in transformers but fails to autonomously offer risk mitigation [7]. Transformer fault classification has been used with the support of the use of support vector machines and genetic algorithms [8]. The scalable data-processing architectures give extra computational insights on processing of monitoring information [9]. Conditions monitoring in renewable-energy systems utilizing sensors has been explored [10]. Deep learning models based on transformers have been developed to detect faults in electric-power-systems as well [11]. Techniques of machine learning that include sweep frequency response analysis aid in fault detection of power-transformers [12]. The latest intelligent transformer condition monitoring systems with AI-based diagnostics provide even more evidence of the opportunities [13].

In spite of the developments, there are often solutions that deal with fault-risk estimation, fault classification and maintenance decision support as individual processes. A comprehensive architecture that links prediction, diagnosis, constrained corrective action and post action evaluation is yet to be developed. This limitation is overcome by the proposed framework which is based on a closed-loop architecture where prediction and diagnosis is used to define the state of the system by which an agentic decision layer will operate. The decision layer, instead of producing just an alarm, transforms the fault probability, diagnosed class, severity, and operating constraints to an admissible corrective action and then updates action priorities based on the observed risk reduction (RR). The study fulfills four contributions: i) a combined transformer fault-management architecture that includes prediction, diagnosis, constrained rectification, and feedback; ii) a context-sensitive expert policy to select corrective action based on the fault probability, diagnosed class, severity and safety constraints; iii) an explicit RR measure of a post action assessment; and iv) reproducible synthetic evaluation with MAE=0.0205, RMSE=0.0271, R^2=0.9881, random forest diagnosis rate of 97.63% with macro F1=0.9762 and simulated scenario-level RR between 65.15% and 78.20%.

2. RELATED WORK

The field of intelligent power-system monitoring has moved beyond traditional machine learning diagnosis to physics-informed models, and adaptive decision support. Physics-informed machine learning has been explored to control stability in smart grids controlled by renewable sources [14]. In other diagnostic applications not related to power-transformers, transformer-based deep-feature learning has been used to diagnose faults in an electronic circuit, demonstrating similar diagnostic procedures [15]. Machine learning has been explored in particular to fault diagnosis of smart transformers [16]. Convolutional and cross-fusion transformer models have been shown to be able to diagnose deep-features of rotating machinery, but the approach is not similar to transformer condition monitoring [17]. The concept of drift-adaptive federated learning has also been considered to predict fault in heterogeneous industrial settings [18]. The presence of online diagnostic methods of identifying faults in the winding of power transformers is another indication of the progress in equipment-specific-monitoring [19].

The other research direction which is key is that of functional integration. Predictive-control strategies have also been explored in enhancing the stability and efficiency of operations in the smart grids of renewable-integrated grids [20]. Predictive maintenance of distribution transformers is also done with the help of machine learning-based predictive maintenance [21]. Such studies enhance smart monitoring and maintenance choices, yet there is a need to explicitly consider the process of fault prediction, fault diagnosis, the selection of corrective actions, and feedback. The current framework thus focuses on a predict-diagnose-action-feedback workflow, which is repeatable with quantitative evaluation of intervention effect and its impact on the priority of actions later on.

Digital-twin methods have been considered in predictive maintenance in smart infrastructure that aids in condition representation and maintenance planning [22]. Digital-twin modeling has been explored in the case of large-scale power transformers, too [23]. Federated learning with privacy has enhanced the correct fault diagnosis of transformers when the data is unbalanced [24]. Software testing automated prioritization techniques give a more comprehensive representation of decision-ranking techniques, but do not specifically relate to transformer maintenance [25].

The developments exhibit increased observability, diagnostic, predictive maintenance, and decision prioritization. However, specific studies tend to focus on single analytic or functional activities, as opposed to how they work together. The proposed structure responds to this integration need by integrating the likelihood of faults, fault-diagnostic classes, severity, safety-constricted by expert valuation, selection of corrective actions and post-corrective actions feedback in a closed-loop transformer fault-management process that can be recreated.

3. SYSTEM MODEL AND METHOD

3.1. Overview of the proposed framework

The proposed architecture combines fault prediction, fault diagnosis, fault rectification planning and feedback in a single sequential decision-making process. Unlike a conventional monitoring pipeline that stops after detection, the present framework forms a system state from the predicted fault probability P_f diagnosed class C, and severity index S, and then selects an admissible action A through the decision policy $h(.)$.

$$X \rightarrow P_f = f_\theta(x) \rightarrow C = g_\varphi(x) \rightarrow A = h(P_f, C, S)$$

Here, X denotes the transformer feature vector, $f_\theta(\cdot)$ is the fault-probability model, $g_\varphi(\cdot)$ is the multi-class diagnosis model, and $h(\cdot)$ is the rectification policy. The main technical distinction between such state-to-action mapping and prediction/diagnosis as separate outputs is the explicit linkage of state to action–the relationship is not implied.

The electrical, thermal, and chemical measurements are preprocessed and provided to the prediction/diagnosis modules as shown in Figure 1. The outputs from these are fused by the agentic decision layer, which incorporates severity and safety restrictions in its issuance of corrective action. This state of the system is then reassessed to determine the extent to which the action either increased or decreased risk.

This ordered interaction gives rise to closed loop interaction: the context in which the action will be taken will be determined by the prediction and diagnosis; the outcome of the action will be measured, and this outcome will determine if the same action should keep the highest level of priority, become more or less important, or be eliminated from the choice of later actions. The effectiveness of actions is assessed with the normalized risk-reduction measure as defined in subsection 3.6.

$$RR = \frac{P_{f,before} - P_{f,after}}{P_{f,before}}$$

where $P_{f,before}$ and $P_{f,after}$ are the estimated fault probabilities before and after rectification, respectively.

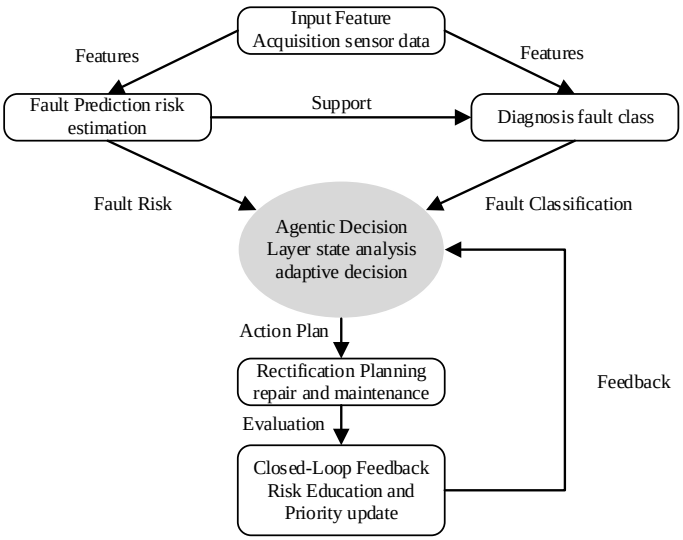


Figure 1. Closed-loop architecture of the proposed agentic AI transformer fault-management framework

3.2. Data representation and feature modeling

There are eight structured variables used to represent the condition of the transformer: load current, voltage, oil temperature, ambient temperature, hydrogen (H_2), methane (CH_4), carbon monoxide (CO), and moisture. These consist of electrical, thermal and chemical variables that are the common feature set in prediction and diagnosis.

$$x = [I, V, T_{oil}, T_{amb}, H_2, CH_4, CO, M]^{\mathrm{T}}$$

The new experiment employs a fully reproducible random number generator (synthetic), with a fixed random seed of 13905. The feature values are sampled from bounded, class-conditioned operating

distributions to enable that the normal, short circuit, thermal, insulation degradation and overload records have different signatures to some degree of overlap. No field data are claimed and the same generated records are used all throughout the reported evaluation.

The data set consists of 5,000 records evenly distributed among the five operating labels: normal, short circuit, thermal fault, insulation degradation and overload. The four fault labels are explicitly given based on the scenario used for the creation of each record. Fault probability is a continuous synthetic target, and is generated by a bounded logistic combination of loading, undervoltage, oil temperature, concentrations of dissolved-gasses, moisture, class offset, and a small Gaussian process variation.

To perform fault probability regression, all 5,000 records are applied. When using a four-class diagnosis, the 1,000 normal records are not included as the diagnosis stage is only activated when a fault-risk condition is detected, and there are 4,000 fault records, 1,000 for each fault class. This separation renders the source, sample size, class distribution and the generation procedure of the labels explicit.

3.3. Fault prediction module

During the prediction stage, the eight-variable feature vector is passed to an extra trees regressor model to predict a continuous fault-risk score. The model trained is based on 500 randomized trees, max_depth is not limited, minLeafSize is equal to 1, max features is equal to 1.0 (all features are used at each split), and random seed is equal to 13905.

$$P_f = f_\theta(x), \ 0 \le P_f \le 1$$

The records are randomly placed into 5,000 records, composed of 80% training (4,000) and 20% held-out for testing (1,000). Five-fold cross-validation is used only on the training partition for stability assessment purposes; all the principal MAE, RMSE and R^2 values provided in section 4, are the values from the set never used for the training. The five-fold training results for extra trees are MAE=0.0204±0.0005, RMSE=0.0268±0.0006, and R^2=0.9884±0.0006.

Both prediction baselines are based on the same feature set, held-out protocol involves a decision tree regressor with maximum depth of 8 and minimum leaf size of 4, and a linearly-engined support-vector regressor with scaled inputs ($C = 1.0$ and $\varepsilon = 0.05$).

3.4. Fault diagnosis module

Random forest classifier is used to find the four types of faults at the diagnosis stage which are short circuit, thermal fault, insulation degradation, and overload. The classifier has been implemented with 300 trees, 12 maximum tree depth, 2 minimum leaf size, square root sampling of features at every split, balanced class weights, and random seed of 13905.

$$C = g_\varphi(x) \in \{SC, TF, ID, OL\}$$

For the diagnosis model, the probabilities of classes are generated by the proportion of trees that vote for each class and the class with the maximum probability is chosen as class C. There are 4,000 records all up in the 4 fault classes. Stratified 80/20 means 3,200 training records and 800 held-out test records–200 test samples per class. Five-fold stratified cross validation is used in the training partition. The evaluation of the systems is done by accuracy, precision, recall, and F1-score.

$$C = argmax_i p(c_i \mid x)$$

3.5. Agentic rectification planning module

The rectification module translates the analysis results into an actionable response. The decision mechanism in the reported experiment is an expert-designed rule-based policy where there are operational and safety constraints, but no reinforcement learning (RL) is used. Action selection is automatic during evaluation, the action definition and the rules to respond to risks are pre-defined for reproducibility.

$$A = h(P_f, C, S)$$

The policy evaluates fault probability P_f, diagnosed class C, and severity S. Slow circuit breaker trips are used for short-circuit, branch isolation, and immediate inspection; slow circuit breaker trips are used for thermal, load reduction and cooling increase; slow circuit breaker trips are used for degradation of insulation, domain generation algorithm (DGA)/insulation maintenance with temporary derating; and overloads are used for load redistribution–derating. These actions are automatically picked up from the diagnosed state depending on the safety constraints.

The pre-action probability is multiplied by a class specific residual-risk factor of 0.22 (short circuit), 0.28 (thermal), 0.35 (insulation degradation) or 0.25 (overload) and bounded by a severity adjustment of -0.05 (S-0.5). Pre- and post-action probabilities are then used in computing the risk-reduction metric. These factors represent expert rule policy; they are not learned by RL.

The achieved decision path is summarized in Figure 2. Prediction, diagnosis and severity are combined in one state, feasible actions derived from the expert rule set, and inadmissible actions eliminated due to operational/safety constraints, and then highest-priority admissible corrective action is carried out. Only post-action RR is used to update the priority of predefined actions, but NOT for RL. The same sequence is also available in procedural form in Algorithm 1, and will be kept for reproducibility.

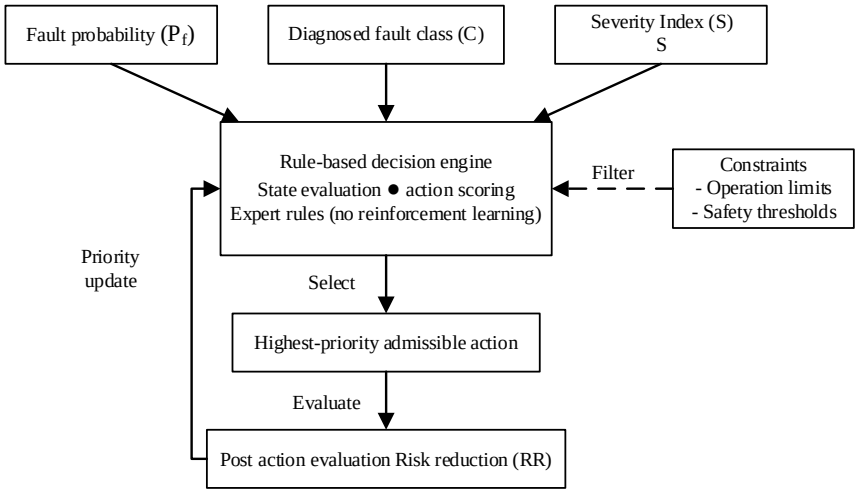


Figure 2. Context-aware agentic decision process used for rectification planning

Algorithm 1. Agentic rectification decision process
Input: predicted fault probability P_f, diagnosed fault class C, severity index S
Output: corrective action A
1: Acquire state (P_f, C, S)
2: Compute severity-adjusted risk
3: Score predefined actions using expert policy h(P_f, C, S)
4: Remove actions that violate operational or safety constraints
5: Execute the highest-priority admissible action A
6: Re-estimate fault probability and compute risk reduction (RR)
7: Update action priority for the same state class using RR
8: Return A

3.6. Closed-loop feedback mechanism

The feedback mechanism evaluates whether the selected action actually reduced predicted risk. After each simulated intervention, the post-action state is passed through the same predictor to obtain $P_{f,after}$, and the normalized RR is computed.

$$RR = \frac{P_{f,before} - P_{f,after}}{P_{f,before}}$$

The direction of RR change: RR>0=improvement, RR=0=no change, RR<0=action increased estimated risk. By normalizing this measure, actions can be compared among scenarios that have varying starting fault probabilities.

It is not about RL it is an update in the practice. The predefined actions change their priority score depending on positive RR and decrease otherwise while safety constraints are unchanged.

$$q_{k+1}(s,a) = q_k(s,a) + \eta \cdot RR$$

where $q_k(s,a)$ is the priority of action a in discretized state s at iteration k and η is a small adaptation factor. The update modifies only the ordering of the predefined set of actions, it does not introduce new actions nor

train a reinforcement-learning policy. This implementation allows to reproduce the closed-loop adaptation with the expert-policy assumption of the present work. The observed state after the action is put back into the prediction stage, the action is checked by RR and the priority of the action is changed accordingly before the next decision step.

3.7. Experimental setup

The quantitative revision experiment is carried out as a repeatable Python-based workflow, instead of a code that cannot be found and is unavailable. The math packages (Python 3.13.5, scikit-learn 1.8.0, NumPy 2.3.5, and pandas 2.2.3) are run on an x86_64 environment with 5 virtual CPUs and about 5.9 GiB of RAM. As illustrated in Figure 3, the rectification process follows a closed-loop feedback mechanism in which the system state is observed, an admissible corrective action is selected under predefined safety constraints, post-action risk is re-estimated, and the RR ratio is evaluated to update action priorities for subsequent corrective interventions. The other points, 5–7, follow from the provided reproducibility code and synthetic data.

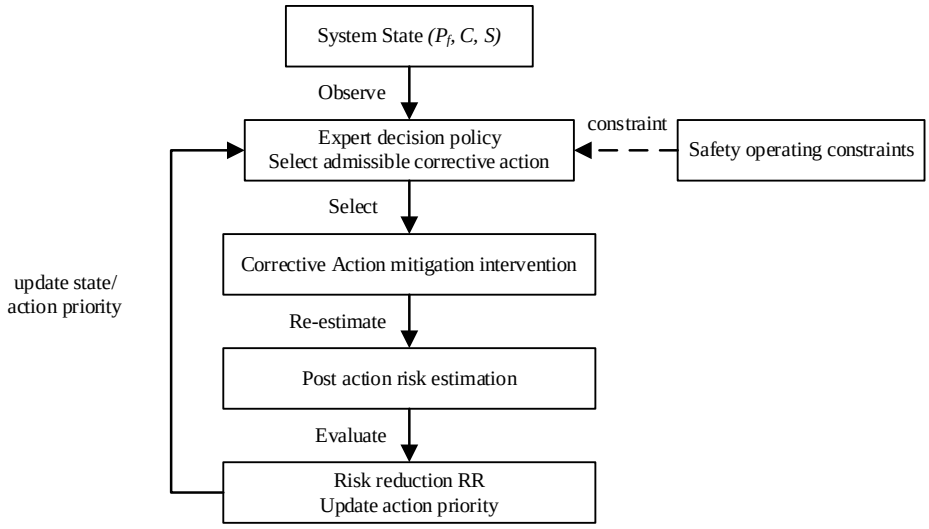


Figure 3. Closed-loop feedback and action-priority update after rectification

Each of the 5,000 records generated by the synthetic generator is balanced, with 1,000 normal, 1,000 short-circuit, 1,000 thermal-fault, 1,000 insulation degradation, and 1,000 overload cases. Load-current, voltage, oil temp, ambient temp, moisture, H_2, CH_4, CO are recorded in each record. The operating/fault label is specified by scenario membership and a bounded logistic stress function specifies the continuous fault-probability target.

A stratified 80:20 split of the total of 5,000 means 4,000 training instances and 1,000 held-out test instances for regression. The 4,000 fault records are then divided into 3,200 training samples and 800 samples for testing and held out of the training set. Five-fold cross validation is performed only on each of the training sets, while the main numerical results in section 4 are the held-out test results.

The extra trees regression (500 trees, max_depth=None, min_samples_leaf=1, max_features=1.0) is used for fault-probability prediction. Random forest classification (300 trees; max_depth=12; min_samples_leaf=2; max_features=sqrt; class_weight=balanced) is used to implement fault diagnosis. Seed used for all randomized procedures is 13905. During the rectification stage there is no use of RL, and the explicit expert rule policy presented in section 3.5 is used.

The prediction model is assessed based on the MAE, RMSE, and R^2; the diagnosis model is assessed based on the accuracy, macro precision, macro recall, macro F1, and the confusion matrix; and the rectification model assessing based on RR. The following code generates the data shown in the following regenerated figures: the reported values are independent in 5-7, using the same fixed values for the seed and parameter.

The regression baselines are a decision tree with max_depth=8 and min_samples_leaf=4, and a standardized linear support vector regression (SVR) with $C = 1.0$ and $\varepsilon = 0.05$. This computational configuration and the actual model settings are now explicitly given to address the reproducibility.

Simulation model of the proposed framework is as shown in Figure 4 and reflects the combined execution of the electrical and thermal dynamics, fault probability estimation, agentic decision-making and

corrective action mechanisms in an integrated environment. The model describes the interplay between the system dynamics and the decision layer and allows the analysis of the closed-loop behavior in different operating conditions.

There is a closed loop signal flow, within the system shown in Figure 4, between the operating variables, risk estimation, decision logic, corrective action, and post-action evaluation. These are quantitative results derived from the reproducibility workflow.

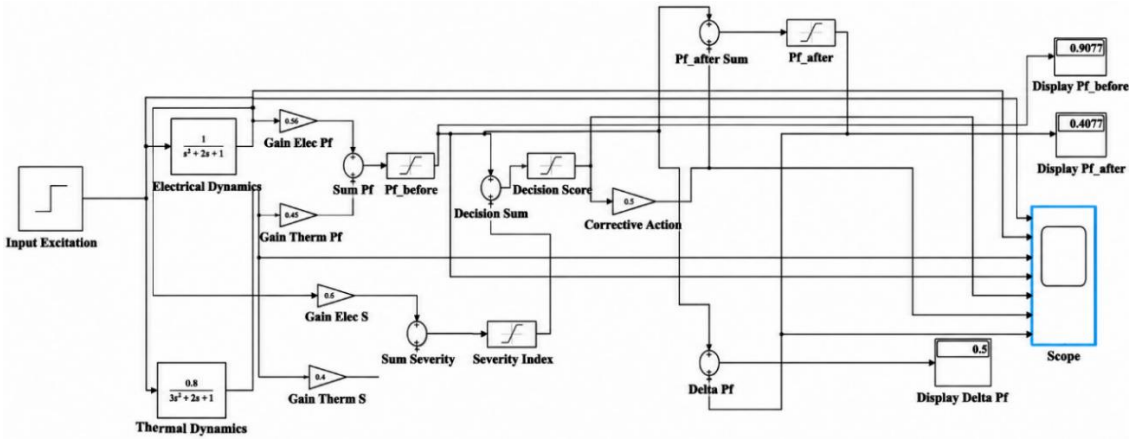


Figure 4. System-level illustration of the proposed closed-loop transformer fault-management workflow

4. RESULTS AND DISCUSSION

4.1. Fault prediction performance

The proposed extra trees predictor gives the best MAE=0.0205, RMSE=0.0271, and R^2=0.9881 in the 1,000-sample held out regression test set. Five-fold cross-validation within the training partition gives MAE=0.0204±0.0005, RMSE=0.0268±0.0006, and R^2=0.9884±0.0006, showing low fold-to-fold variation.

Using the same held-out protocol, the decision tree baseline gives MAE=0.0400, RMSE=0.0543, and R^2=0.9523, while linear SVR gives MAE=0.0227, RMSE=0.0287, and R^2=0.9867. Therefore, extra trees proved to be the best compromise between lower MAE/RMSE and R^2 among the tested predictors as revealed in Figure 5.

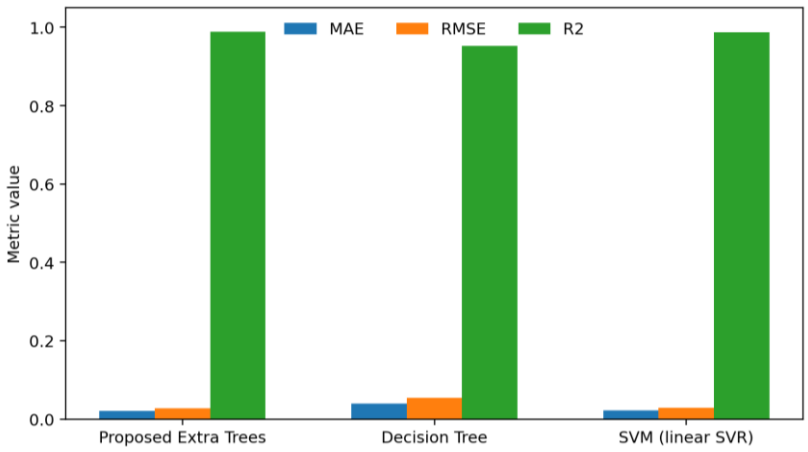


Figure 5. Held-out fault-probability regression performance of extra trees, decision tree, and linear SVR

Figure 5 isolates the prediction-model effect since they share the same eight variables and constitute a held-out protocol. Improvements of the extra trees over linear SVR are relatively small, and its improvements over the single decision tree are more significant; but the key contributions of the principals to the engineering are the integration of risk prediction with diagnosis, constraints on action selection, and feedback, downstream.

4.2. Fault diagnosis performance

The diagnosis stage of the random forest model has an accuracy of 97.63%, a macro precision of 0.9762, a macro recall of 0.9763 and a macro F1 of 0.9762 on the 800-sample held-out fault test set. Five-fold training cross-validation gives 96.62%±0.70% accuracy and macro F1=0.9662±0.0071. The class scores are 0.9648, 0.9851, 1.0000 and 0.9550 for short-circuit, thermal fault, insulation degradation and overload respectively.

On average, 192/200 short circuit, 198/200 thermal, 200/200 insulation-degradation and 191/200 overload test cases are correctly classified as shown in Figure 6. The most frequent error ranges are between short circuit and overload or between overload and thermal, and these error ranges partly overlap. The class-level result is important because the rectification module is based on C directly: the overall fault probability may be high, but the class level may not, which means that people might select the wrong intervention.

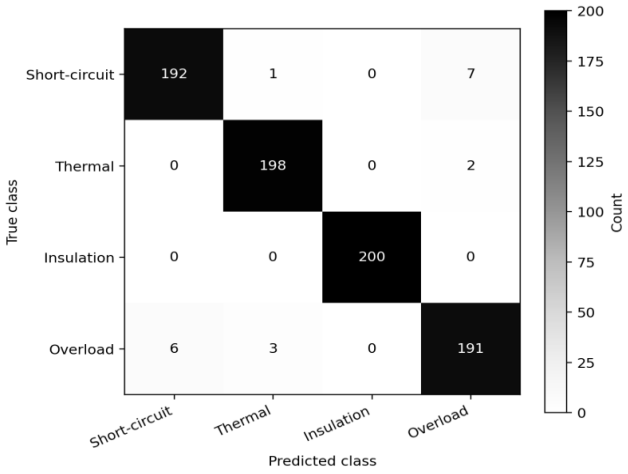


Figure 6. Held-out confusion matrix for random forest four-class transformer fault diagnosis

4.3. Rectification effectiveness

The explicit expert rule policy is tested on an independent set of 800 fault cases from the 1024 cases not used to train the rectification methods. The average of this mean RR is 78.20% for short circuit, 72.12% for thermal fault, 65.15% for insulation degradation and 75.21% for overload.

The mean pre-action/post-action fault probabilities for short-circuit, thermal fault, insulation degradation and overload are shown in Figure 7, respectively 0.6747/0.1455, 0.6305/0.1743, 0.5993/0.2075, and 0.6404/0.1576. The policy and residual-risk factors are expert rules that are predefined, and the selection of action is automatic from the state.

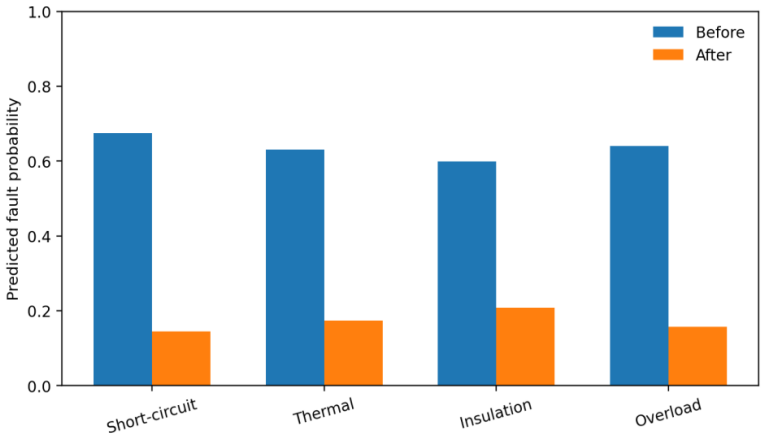


Figure 7. Mean fault probability before and after expert rule-based rectification across four fault scenarios

The reduction in all scenarios confirms the purpose of the closed loop; predicting and diagnosing is followed by action, which is followed by reevaluation of the post-action state, which is then fed back to determine priority for action.

4.4. Discussion

The remediated outcomes back the desired function of the proposed framework, which is not just as a prediction model, but as a holistic fault management workflow. For the held-out regression set, extra trees achieved MAE=0.0205, RMSE=0.0271, and R^2=0.9881 while the random forest diagnosis stage reported an accuracy of 97.63% and a macro F1 of 0.9762. The results show that under the controlled synthetic evaluation, the two analytical stages can provide a consistent description of the states of the downstream decision layer. The primary engineering benefit comes not from the numerical difference between the strong predictors, these simply present a risk diagnosis; but from the clear incorporation of the risk-diagnosis and fault class into a set of constrained corrective actions, and ensuing post-action reassessment.

The regression comparison also makes sense the prediction stage contribution. The linear SVR model prediction error is already close to extra trees model while the decision tree baseline model has the large visibly prediction error. Thus, the reported framework is not unique with respect to the regressor. Instead, the predictor is part of a closed loop system. The prediction module can be combined with the diagnostic and rule-based decision module and the chosen extra trees configuration has the lowest MAE and RMSE compared to the other tested configurations. This interpretation helps prevent the misinterpretation of the entire benefit of the framework to marginal differences in model performance from regression.

The diagnosis results are particularly interesting as the corrective rule depends on the class. The confusion matrix indicates that 192/200, 198/200, 200/200, and 191/200 short circuit, thermal fault, insulation degradation and overload cases, respectively, are correctly classified in the held-out test set. The rest of the errors are predominantly between short-circuit and overload or overload and thermal (partial overlap in synthetic operating signatures), which is consistent. Confusion matters more in a closed loop maintenance scenario as the wrong class may result in the wrong intervention. This is why the framework is divided into two stages, continuous risk estimation, and multi-class diagnosis, and both results are utilized in the building of the decision state.

The rectification results show that the aim of the agentic layer is achieved, but not necessarily to regard the learning as RL or as full autonomy. The overall average RR obtained with the predefined expert policy were 72.67% (mean value): 78.20% for short circuit, 72.12% for thermal fault, 65.15% for insulation degradation and 75.21% for overload. The lower reduction due to insulation loss has also been similar to the higher residual-risk factor assigned to the class by the expert policy. Thus, differences between fault scenarios are to be taken as a result of the rule set and the severity adjustment, which is specified, and not as the result of learning an independent control behavior. The action-selection mechanism is autonomous only in the sense that the most suitable action that is admissible at the given state and constraints is automatically selected among a limited set of predefined actions after the evaluation of the state and constraints.

Another benefit of the proposed organization is the post action assessment that is explicitly stated. The present workflow re-estimates the probability of fault after a simulated intervention, while conventional condition-monitoring techniques end upon detection or diagnosis, the change in fault probability is represented using the normalized risk-reduction measure. This feedback signal doesn't modify the control policy, but it only changes the priority ordering of the action sets that the expert rule base allows. This is crucial for reproducibility and safety, since the available actions and operating restrictions do not change. The feedback mechanism then offers a quantifiable means of determining the effectiveness of an intervention retaining an open and visible decision pathway.

The same fixed random seed is used for all runs reported, explicit model hyperparameters are used, the classes are balanced, a held-out set for testing is used, and training-only five-fold cross-validation is employed, which increases the level of reported experiment reproducibility. The near-identical cross-validation statistics and held-out extra trees statistics suggest little variation among the data generated. However, this does not mean that this is the same as reproducibility under a synthetic generator, as it is actually a different type of validity. The data distributions, class offsets, residual-risk factors, and severity adjustments were developed for this study and are not intended to represent every transformer aging mechanism, every sensor error, every maintenance history or every operating environment found in utility data. The outcome should be considered as a check for the computational procedure and the internal consistency of the closed loop and not as a sign of reliability of the field.

Several considerations, however, are not included in the evaluation here but are important for its practical deployment: the unbalanced class frequencies, the absence of dissolved-gas measurements, delayed laboratory results, sensor drift, seasonal loading variation and the gradual development of fault signatures can all be found in real transformers. Furthermore, coordination with protection schemes and maintenance

procedures would be needed for a field implementation, ensuring that any action that is automatically recommended does not overrule established safety rules. These factors are not a contra-indication to the current proof-of-concept, but they are conditions that need to be explored before the framework can be used as a suitable framework for operational decision support.

Thus, the validation of the future should be done in steps. Firstly, the prediction and diagnosis modules should be tested with operational or publicly available fault records of transformer, with naturally unbalanced fault classes and missing data conditions. Second, hardware-in-the-loop or real-time simulation may be employed for evaluating latency, timings of actions and interaction with protection constraints. Third, the expert rule policy can be compared with learning-based or optimization-based decision strategies with the same set of admissible actions and safety constraints. Explainable-AI mechanisms would also help in displaying the causes of a risk prediction or diagnosis, and the reasons a specific corrective action was given the highest priority. These extensions would constitute a better foundation for evaluating generalization, safety and operationally usefulness and maintain the clear closed loop structure shown here.

5. CONCLUSION

This work proposed a closed-loop agentic AI system to handle transformer fault management, which integrates fault-probability prediction, multi-class diagnosis, constraint-aware rectification planning and post-action feedback into one reproducible workflow. The design is deliberately structured to distinguish data-driven prediction of fault risk from the expert-derived action policy: prediction of fault risk is done using extra trees, the fault diagnosis is done using random forest and corrective actions are selected from a prespecified admissible set (operational and safety constraints). Post-action state is then re-evaluated and the impact of the intervention chosen is measured with normalized RR.

The held-out regression set results under the fixed synthetic evaluation protocol are: MAE=0.0205, RMSE=0.0271, and R^2=0.9881 for extra trees predictor. During the diagnosis stage, 97.63% is the accuracy and macro F1=0.9762 on held-out fault set. The estimated reduction in risk by the expert rule policy was on average 72.67% (range of 65.15–78.20%) at a scenario level. These results make it clear that prediction and diagnosis services can be passed to a repeatable action-selection and feedback loop with the assumptions made in the data generation process and the assumed policy.

Hence principal contribution is in the integration of prediction, diagnosis, constrained decision making and evaluation of action and not in the claim of autonomous, unconstrained decision making. It is not a RL framework, does not propose new corrective actions and does not make field deployment claims. It can't even add new actions to its list; it can only select and reprioritize from a pre-defined list based on the diagnosed state, severity, safety constraints and observed post-action RR. This "bounded formulation" brings out the decision pathway clearly, and keeps the simulated maintenance process interpretable.

The current results are achieved when using synthetic operating profiles and expert determined residual-risk factors. They do show reproducibility of the proposed workflow and internal consistency of this workflow but cannot prove the reliability across real transformer fleets, different sensor configurations and practical protection environments. Future research will thus be on one hand concerned with the validation by the data of the operational transformers, on the other hand with naturally imbalanced fault conditions, with hardware in the loop testing and with direct comparison with the learning-based decision strategies. Ensuring safety and traceability will also be crucial to the overall development of the framework into a practical decision support system, with the integration of explainability and uncertainty handling.

ACKNOWLEDGMENTS

The authors would like to thank their respective institutions for providing the academic support and resources necessary to carry out this research.

FUNDING INFORMATION

This research received no external funding.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
P. Lakshmi Narayana	✓	✓	✓	✓		✓			✓					
Robbi Rahim		✓		✓	✓					✓				
Kavitha Manoharan	✓			✓						✓		✓	✓	
A. V. Prabu				✓	✓					✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known financial or non-financial conflicts of interest that could have influenced the research presented in this manuscript.

DATA AVAILABILITY

The data supporting the findings of this study were generated synthetically using electrical, thermal, and dissolved-gas parameters. A total of 5,000 synthetic transformer operating records were generated and analyzed for model training, validation, and performance evaluation. The synthetic dataset and associated data-generation details are available from the corresponding author upon reasonable request.

REFERENCES

[1] N. K. Sharma, P. K. Tiwari, and Y. R. Sood, "Review of Artificial Intelligence Techniques Application to Dissolved Gas Analysis on Power Transformer," *International Journal of Computer and Electrical Engineering*, vol. 3, no. 4, pp. 577–582, 2011, doi: 10.7763/ijcee.2011.v3.383.

[2] J. Cui, W. Kuang, K. Geng, and P. Jiao, "Intelligent fault diagnosis and operation condition monitoring of transformer based on multi-source data fusion and mining," *Scientific Reports*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-91862-8.

[3] A. Abu-Siada and S. Islam, "A new approach to identify power transformer criticality and asset management decision based on dissolved gas-in-oil analysis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 19, no. 3, pp. 1007–1012, 2012, doi: 10.1109/TDEI.2012.6215106.

[4] J. Dai, H. Song, G. Sheng, and X. Jiang, "Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 24, no. 5, pp. 2828–2835, 2017, doi: 10.1109/TDEI.2017.006727.

[5] C. Rajan, "Congestion- and fault-severity-aware escape routing with local recovery for multicore networks-on-chip," *VSF Journal of Electronics and Communication Engineering*, vol. 1, no. 1, pp. 30–37, 2026.

[6] O. J. M. Smith and F. de Mendonca, "AI-driven predictive maintenance framework for fault detection and performance optimization in smart grid and renewable energy systems," *National Journal of Electrical Electronics and Automation Technologies*, vol. 1, no. 2, pp. 42–49, 2025, doi: 10.17051/JEEAT/01.02.06.

[7] M. Duval, "A review of faults detectable by gas-in-oil analysis in transformers," *IEEE Electrical Insulation Magazine*, vol. 18, no. 3, pp. 8–17, 2002, doi: 10.1109/MEI.2002.1014963.

[8] T. Songjiang, G. W. Kari, and Z. Ziwei, "Transformer fault diagnosis based on support vector machine and genetic algorithm [J]," *Journal of Tsinghua University (Natural Science Edition)*, vol. 58, no. 7, pp. 623–629, 2018.

[9] S. R. K. Reddy, "A fault-tolerant scalable data engineering architecture for real-time streaming analytics in distributed systems," *Journal of Scalable Data Engineering and Intelligent Computing*, vol. 3, no. 3, pp. 1–10, 2026.

[10] F. P. G. Márquez and I. S. Ramírez, "Condition monitoring system for solar power plants with radiometric and thermographic sensors embedded in unmanned aerial vehicles," *Measurement: Journal of the International Measurement Confederation*, vol. 139, pp. 152–162, 2019, doi: 10.1016/j.measurement.2019.02.045.

[11] D. H. Yoon and J. Yoon, "Development of a real-time fault detection method for electric power system via transformer-based deep learning model," *International Journal of Electrical Power and Energy Systems*, vol. 159, 2024, doi: 10.1016/j.ijepes.2024.110069.

[12] M. Bjelić, B. Brković, M. Žarković, and T. Miljković, "Machine learning for power transformer SFRA based fault detection," *International Journal of Electrical Power and Energy Systems*, vol. 156, 2024, doi: 10.1016/j.ijepes.2023.109779.

[13] A. Revera and M. Abbas, "Condition Monitoring of Power Transformers Using AI-Driven Diagnostic Systems," *National Journal of Electrical Machines and Power Conversion*, vol. 1, no. 3, pp. 1–8, 2025.

[14] D. Abdullah, "Physics-informed machine learning for real-time stability control in renewable-dominant smart grids," *National Journal of Renewable Energy Systems and Innovation*, vol. 2, no. 1, pp. 1–9, 2026.

[15] I. Z. Mukti and E. R. Khan, "Real-Time Streaming Analytics for Predictive Fault Diagnosis in Cyber-Physical Energy Infrastructures," *Progress in Electronics and Communication Engineering*, vol. 3, no. 2, pp. 15-22, 2026, doi: 10.31838/ECE/03.02.03.

[16] T. M. S. Kumar, "Fault Diagnosis in Smart Transformers Using Machine Learning Techniques," *Journal of Reconfigurable Hardware Architectures and Embedded Systems*, vol. 2, no. 2, pp. 16–23, 2025, doi: 10.17051/JRHAES/02.02.03.

[17] T. Lin, Y. Zhu, Z. Ren, K. Huang, and D. Gao, "CCFT: The Convolution and Cross-Fusion Transformer for Fault Diagnosis of Bearings," *IEEE/ASME Transactions on Mechatronics*, vol. 29, no. 3, pp. 2161–2172, Jun. 2024, doi: 10.1109/TMECH.2023.3312935.

[18] C. Van, M. H. Trinh, and T. Shimada, "Drift-adaptive asynchronous federated learning for fault prognosis in heterogeneous

industrial plants," *Transactions on Internet Security, Cloud Services, and Distributed Applications*, vol. 3, no. 3, pp. 23–30, 2026.
[19] A. Abu-Siada and S. Islam, "A novel online technique to detect power transformer winding faults," *IEEE Transactions on Power Delivery*, vol. 27, no. 2, pp. 849–857, 2012, doi: 10.1109/TPWRD.2011.2180932.
[20] S. Shaik and R. G. Luedke, "AI-Driven Predictive Control for Renewable-Integrated Smart Grids: Toward Enhanced Stability and Efficiency," *National Journal of Intelligent Power Systems and Technology*, vol. 1, no. 2, pp. 44–51, 2025, doi: 10.17051/NJIPST/01.02.06.
[21] V. Biradar, D. Kakeri, and A. Agasti, "Machine Learning based Predictive Maintenance in Distribution Transformers," in *2024 8th International Conference on Computing, Communication, Control and Automation, ICCUBEA 2024*, 2024, pp. 1–5, doi: 10.1109/ICCUBEA61740.2024.10774993.
[22] C. A. Prasath and V. T., "AI-Enabled Digital Twin Framework for Predictive Maintenance in Smart Urban Infrastructure," *Journal of Smart Infrastructure and Environmental Sustainability*, vol. 2, no. 1, pp. 1–10, 2025, doi: 10.17051/JSIES/02.01.01.
[23] S. Veerappan, "Digital Twin Modeling for Predictive Maintenance in Large-Scale Power Transformers," *National Journal of Electrical Machines & Power Conversion*, vol. 1, no. 1, pp. 39–44, 2025.
[24] Q. Wu, C. Dong, F. Guo, L. Wang, X. Wu, and C. Wen, "Privacy-Preserving Federated Learning for Power Transformer Fault Diagnosis With Unbalanced Data," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 4, pp. 5383–5394, 2024, doi: 10.1109/TII.2023.3333914.
[25] N. Jagadhabi, "Machine Learning Models for Automated Testing Prioritization in Continuous Integration Systems," *SECITS Journal of Scalable Distributed Computing and Pipeline Automation*, vol. 3, no. 1, pp. 58–62, 2025.

BIOGRAPHIES OF AUTHORS



P. Lakshmi Narayana is an assistant professor in the Department of Electrical and Electronics Engineering at Vignan's Foundation for Science, Technology and Research (Deemed to be University), AP, India. His academic and research focus spans power electronics, electrical machines, smart grid integration, and control systems engineering, contributing actively to both classroom instruction and technical research within the institution. He can be contacted at email: pln_eee@vignan.ac.in.



Robbi Rahim is a highly accomplished academic researcher and data scientist based in Medan, Indonesia, with a robust publishing portfolio spanning data science, artificial intelligence, information security, and renewable energy systems. With over 140 total publications and 850 citations, his work demonstrates significant scientific impact across diverse domains, particularly in algorithmic optimization, decision support systems, cryptographic models, and machine learning applications. In addition to his primary research, he actively contributes to the broader scientific community through extensive peer review, editorial collaboration, and advancing knowledge dissemination within international academic networks. He can be contacted at email: usurobbi85@zoho.com.



Kavitha Manoharan is a faculty member in the Department of Electronics and Communication Engineering (ECE) at the Saveetha School of Engineering, Saveetha University, Chennai, India. Her academic and research focus lies in electronic systems and communication technologies. She is dedicated to advancing engineering education and technical research within the Saveetha academic ecosystem. She can be contacted at email: mkseee2007@outlook.com or kavithamece@gmail.com.



Dr. A. V. Prabu is an associate professor in the Department of Electronics and Communication Engineering at Koneru Lakshmaiah Education Foundation (Deemed to be University), Andhra Pradesh, India. He holds a B.E. from Anna University, an M.Tech. from Biju Patnaik University of Technology (BPUT), and a Ph.D. from Jawaharlal Nehru Technological University Kakinada (JNTUK). With over 20 years of academic experience, his research interests include wireless sensor networks, internet of things (IoT), wireless communications, smart healthcare, embedded systems, and AI-enabled intelligent sensing. He can be contacted at email: prabu.deva@kluniversity.in.