Polytechnic Journal
Polytechnic Journal

Volume 15 | Issue 1 Article 7

Detection of Electrical Fires in Residential Buildings Using a Gradient Boosting Machine Algorithm

Sharaban Taha Ahmed
Department of Electrical Engineering, Salahaddin University-Erbil, Erbil, Iraq, sharaban.ahmed@su.edu.krd
Fadhil Aula
Department of Electrical Engineering, Salahaddin University-Erbil, Erbil, Iraq

How to Cite This Article
Ahmed, Sharaban Taha and Aula, Fadhil (2025) "Detection of Electrical Fires in Residential Buildings Using a Gradient Boosting Machine Algorithm," *Polytechnic Journal*: Vol. 15: Iss. 1, Article 7.
DOI: https://doi.org/10.59341/2707-7799.1860

Detection of Electrical Fires in Residential Buildings Using a Gradient Boosting Machine Algorithm

Data Availability Statement

The data supporting the findings of this study are publicly available and are included within this published article.

ORIGINAL ARTICLE

Detection of Electrical Fires in Residential Buildings Using a Gradient Boosting Machine Algorithm

Sharaban Taha Ahmed*, Fadhil Aula

Department of Electrical Engineering, College of Engineering, Salahaddin University-Erbil, Erbil, Iraq

Abstract

The paper discusses a technique for detecting electrical fires in residential buildings using the Gradient Boosting Machine (GBM) algorithm. The features of the algorithm process data related to current, voltage, and total harmonic distortion (THD) from electrical systems, considering resistive loads, such as heating appliances, and inductive loads, like refrigerators and washing machines. The technique underscores the relationship between electrical characteristics and fire risks, demonstrating that the gradient boosting machine can accurately predict fire hazards under various fault conditions, including arc faults, overvoltage, and contact opening. Results from MATLAB simulations confirm the algorithm's efficacy and high accuracy rates for heating systems and induction motors across different fault types that could lead to electrical fires in buildings. These results highlight the significance of effective feature selection in enhancing the algorithm's performance while addressing some imprecision, particularly regarding the two different load types. Ultimately, the Gradient Boosting Machine represents a promising approach to improving the safety of electrical systems and supporting fire detection strategies.

Keywords: Gradient boosting machine (GBM), Total harmonic distortion (THD), Electrical fires, Arc faults, Machine learning

1. Introduction

Fire safety is becoming a major concern due to the advancements in daily living. Fire dangers are devastating in terms of human life and extremely dangerous and demeaning regarding home and company safety measures. Responding to these dangerous cases as soon as feasible is the clear method to reduce the type of fire [1]. Traditional fault detection methods for arc faults involve time domain analysis, frequency domain analysis, or physical detection. These methods convert electrical impulses into useful data. AI-based methods use machine learning algorithms to analyze power system data to identify fault issues, offering a more accurate and efficient alternative [2]. Aziz *et al.* [3] studied machine failure prediction using data-driven predictive maintenance, a machine learning (ML) technique that identifies potential system malfunctions and alerts when prone to failure. The author introduced Artificial Intelligence Monitoring, a unique framework that anticipates and assesses equipment state, and explained the gradient boosting machine (GBM) algorithm. GBM is preferred due to its simplicity, interpretability, and robustness with small datasets. It is also compatible with legacy systems and allows for customization. Its level-wise tree growth is easier to understand and debug compared to LightGBM's more advanced regularization techniques. Furthermore, GBM provides greater flexibility in customizing the boosting process, making it suitable for specific applications. In contrast, XGBoost's complexity and longer training times may not be ideal for more straightforward tasks or smaller datasets [4]. GBMs demonstrated excellent accuracy and generalization in real-world applications, providing insights into model design for deeper analysis and investigation of modeled effects [5].

In the context of supervised machine learning, as Markoulidakis *et al.* [6]. explained, the challenge of

Received 30 January 2025; accepted 18 May 2025.
Available online 16 June 2025

* Corresponding author.
E-mail address: sharaban.ahmed@su.edu.krd (S.T. Ahmed).

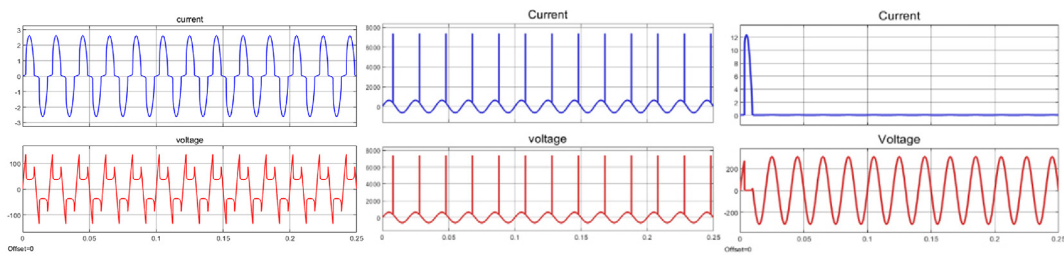
https://doi.org/10.59341/2707-7799.1860

data classification refers to an algorithm's capacity to predict a sample's class label based on a collection of factors known as features or explanatory variables. Two steps make up the solution for data classification, which is an example of pattern recognition: (a) a training phase in which the exploited machine learning algorithm is trained to use a set of training data, that is, data for which the corresponding class label and feature values are known, and (b) applying the trained machine learning algorithm to datasets with unknown class labels and known feature values to predict the class labels. Andrea *et al*. [7] introduced a new mathematical representation of an arc in a circuit, incorporating arc temperature, voltage, and current inertia. The author detailed the equations for modeling small signal behavior, likely supporting MATLAB analysis and simulations. The mathematical and simulation models should account for arc voltage and current relationships, time constants, and response dynamics while integrating the existing models for improved applicability and accuracy. Han *et al*. [8] proposed a low-voltage series arc fault detection method using a parallel Alex Net deep learning network. This method directly detects current signals under normal operation and series arc fault conditions, extracting hidden characteristics from current signal data. Gao *et al*. [9] developed Arc Net, a CNN-based arc detection model, with a 99.47 % accuracy rate. The model was implemented on a Raspberry Pi 3B, which can detect arc faults and identify load types. The model's feasibility in real-time is demonstrated with future work focusing on online learning and unknown load types. Jiang *et al*. [10] said that the adaptive arc detection algorithm, based on machine learning, can be enhanced by using various classifiers, including Support Vector Machine (SVM), Least Squares Support Vector Machine (LSSVM), Artificial Neural Network (ANN), and Convolutional Neural Networks (CNN), but faces challenges in practical application. Han *et al*. [11] presented a recognition method using kernel PCA and firefly algorithm-optimized support vector machine to accurately identify arc faults in voltage and current signals, addressing harmonic interference issues. Wang *et al*. [12] developed a novel framework for identifying series-type low-voltage AC fault arcs. This framework integrates a Markov transfer field, multi-feature fusion, and an improved residual neural network, achieving a recognition accuracy of 99.88 %. Machine learning models for the early detection of electrical fires in residential settings face trade-offs. High-accuracy models may require advanced sensors and computational resources, increasing costs. Complex systems may require professional installation and maintenance, adding to costs. Simpler systems are more user-friendly but may not detect fires as early. Accurate models require extensive datasets, which can be expensive.

This paper introduces a novel technique for detecting three types of faults that can lead to electrical fires, utilizing a single algorithm based on the GBM method. This approach presents unique challenges. As the complexity of identifying common causes increases, understanding the relationship between voltage and current across different load types becomes more difficult. Identifying the common causes and understanding the relationship between voltage and current across various load types becomes increasingly difficult. This paper also examines the effectiveness of GBM in analyzing fire behavior in electrical circuits through fault detection and conducts load tests with a heater and an induction motor. The focus is on selecting voltage, current, and total harmonic distortion (THD) features. This is due to the characteristic variations between voltage and current, which lead to phase shifts (lead and lag) and the production of harmonics in steady-state conditions. MATLAB simulation is used to detect and analyze faults under two different loads, and the machine learning results present the accuracy of the GBM algorithm for each fault.

2. Methodology

Developing fault scenarios in MATLAB Simulink to evaluate residential building loads, including induction motors and heating systems. To achieve this, it is crucial to check and identify each type of fault using the GBM algorithm to ensure accuracy. The process involves recording normal operations and saving the data using outcomes in MATLAB Simulink. Once the data is collected, a fault is applied at a predetermined time limit. Analyzing the differences in current, THD, and voltage as feature selection can effectively detect faults that lead to fires in residential buildings approved by machine learning (ML) results and a confusion matrix (accuracy, precision, recall, and F1 score). This study specifically focuses on three types of faults that can potentially cause fires in residential buildings: arc faults, overvoltage, and contact opening failures. An arc fault is an electrical issue when insulation breaks down, causing electricity to jump through the air across a gap and create an electric arc. This phenomenon can generate high temperatures, potentially leading to fires or damage to electrical components (refer to Fig. 1a). Overvoltage arises when the voltage exceeds normal



(a) (b) (c)

Fig. 1. Three types of faults: (a) arc fault, (b) Overvoltage fault, (c) Contact opening fault.

levels, which can lead to arc discharge and possible damage to equipment (see Fig. 1b). An open contact refers to a situation in electrical circuits where a switch or relay has been disengaged, interrupting the flow of current, in this state, the connection that allows electrical current to pass is broken, which can result in an arc discharge, especially if there is a significant voltage across the open contacts. When an open contact occurs, particularly in devices such as relays or switches, there may be a momentary increase in voltage as the contacts separate. This can lead to the formation of an electric arc, a luminous discharge of electricity that occurs across a gap. Arcing generates intense heat, and if combustible materials are present nearby, it can ignite them, creating a fire hazard (see Fig. 1c). Each of these faults impacts the relationship between voltage and current in various ways, as described by Andrea *et al*. [7].

The GBM algorithm uses specific settings for training and predicting data. Here are the key points and hyperparameters that are used:

- Tree Template ('MaxNumSplits', 20): This setting describes the type of decision tree used in the ensemble. 'MaxNumSplits' limits the number of splits or decision points in each tree, which helps control how complex the trees can get. In this case, it's set to 20.
- Method: 'Bag': This means the algorithm uses Bagging (Bootstrap Aggregating) as its learning technique. Bagging trains multiple learners on different samples of the data to reduce mistakes and improve predictions.
- Learner Type: The base learners are decision trees, as indicated by the Tree template.
- Input Features: The training data includes current and THD values and voltage values for some cases, and faults are the target labels.

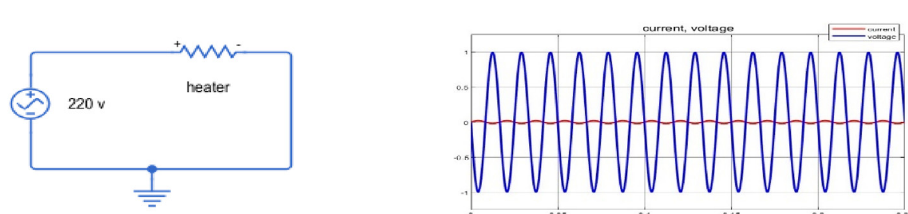
To minimize false positives and false negatives in a classification model, adjust the decision threshold, optimize hyperparameters, improve feature engineering, handle class imbalances, and use cross-validation to ensure the model generalizes well and doesn't overfit the training data.

2.1. Faults in the heating system

The heating system plays a crucial role in the operation of contemporary urban areas, influencing a wide range of activities in both industrial production and everyday home life [13]. Under normal operating conditions, this system behaves like a resistive load, with both voltage and current in phase, as detailed in Fig. 2.

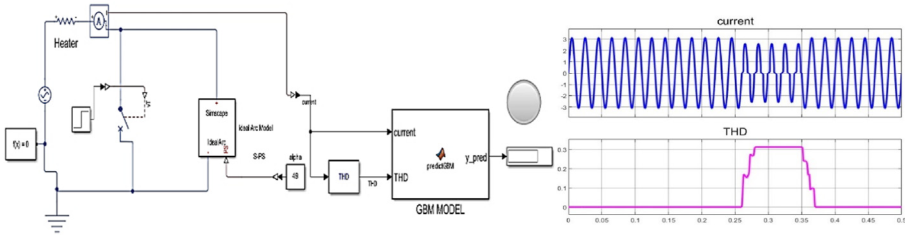
2.1.1. Detecting arc fault in a heating system

The arc fault series is connected to the heater load in the simulation model shown in Fig. 3a. When faults occur from 0.25 to 0.35 s (simulation operation timeframe), current and THD variations can be



(a) (b)

Fig. 2. Heating system: (a) circuit diagram, (b) voltage−current sine wave curve.



(a) (b)

Fig. 3. MATLAB Simulation model arc fault in a heater: (a) circuit diagram, (b) effect of arc fault on current and THD.

observed, as shown in Fig. 3 b. These characteristics are instrumental in creating an effective algorithm utilizing GBM as a machine learning tool. Table 1 displays the results obtained from the GBM and its accuracy metrics.

2.1.2. Detecting contacting open fault in the heating system

Here, the reliance on the current effect and its THD is useful for detecting open contact faults, as shown in Fig. 4, and the machine learning metrics are illustrated in Table 2.

2.1.3. Detecting overvoltage fault in heating system

Fig. 5 illustrates that during this fault, both the current and THD values begin to change in comparison to normal operation, and the GBM algorithm shows remarkable accuracy in detecting overvoltage faults, as indicated in Table 3.

2.2. Faults in induction motors

Priyanka Ray [14] outlined how single-phase motors are used in fans, refrigerators, vacuum cleaners, washing machines, a variety of kitchen appliances, blowers, small agricultural appliances, and other devices because easy to build, dependable, affordable, and easily repairable. Most motors with modest ratings are designed to run on single-phase AC power at standard frequencies. In this simulation, a single-phase induction motor (see Fig. 6a) serves as a suitable representation of the motor characteristics found in a residential building. This type of motor has the same phase characteristic as the inductive load, as illustrated in Fig. 6 b.

Table 1. Machine learning result (Arc fault in heater).

Accuracy	Precision	Recall	F1 score
98.15 %	98.42 %	97.85 %	98.14 %

Table 2. Machine learning result (Contacting open fault in the heater).

Accuracy	Precision	Recall	F1 score
98.81 %	98.41 %	99.21 %	98.81 %

2.2.1. Detecting arc fault in an induction motor

Initially, checking the motor's speed and the RMS values of voltage and current is essential, as illustrated in Fig. 7a. This step is important before identifying the features that will be most beneficial for ML outcomes related to detecting the rate of change. Fig. 7 b illustrates that the arc fault occurred between 0.9 s and 1.2 min, assuming the motor passed the transient state of operation and operated in a steady state. The current and THD proved to be an excellent choice for achieving high accuracy in the GBM algorithm, which achieves an accuracy of 98.39 %, as shown in Table 4.

To verify and improve accuracy, the voltage was selected as an additional feature, as shown in Fig. 8. The arc fault occurs from 0.9 s to 1.2 min; there is a significant change in voltage values for detecting

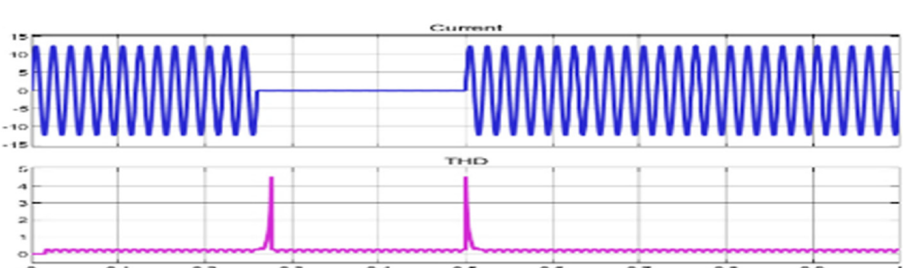


Fig. 4. Open contact fault, the effect on the current and THD in the heater.

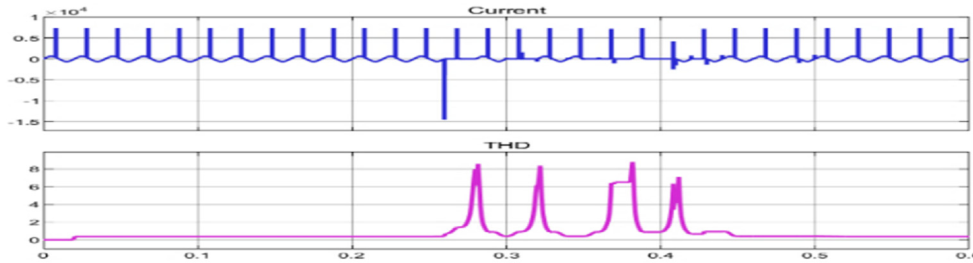


Fig. 5. An overvoltage fault, an effect on the current and THD in the heater.

Table 3. Machine learning result (Overvoltage fault in the heater).

Accuracy	Precision	Recall	F1 score
99.53 %	99.18 %	99.89 %	99.53 %

faults, but the accuracy of the algorithm remains at 98.37 % as shown in Table 5. Detecting this type of fault in induction motors relies solely on current and THD, which are sufficient for achieving high accuracy.

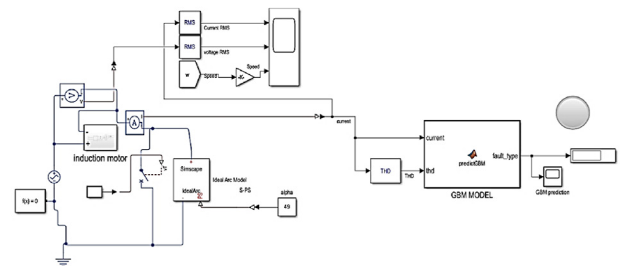
2.2.2. Detecting open contact fault in induction motor

The GBM algorithm for detecting contacting open faults in induction motors achieves an impressive accuracy of 98.65 %, as presented in Table 6. This was illustrated using MATLAB Simulink, as shown in Fig. 9a. Choosing the current and its THD as selection features is a logical decision, as shown in Fig. 9 b. These two features are essential for detecting open contact faults that occur between 0.25 and 0.5 s in this Simulink test.

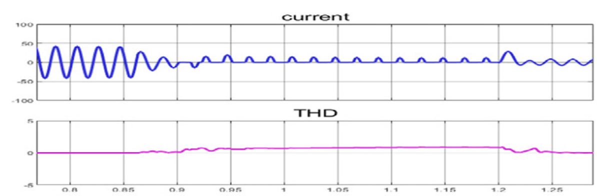
2.2.3. Detecting overvoltage fault in induction motor

The overvoltage fault occurred from 0.5 to 0.7 s, as shown in Fig. 10. During this period, the current amplitude decreased, and THD showed significant harmonic changes. Predicting the GBM algorithm by THD and current feature selection achieves an accuracy of 96.24 %, as preferred in Table 7.

The GBM algorithm, capable of predicting overvoltage using current and THD, reached approximately the same accuracy (refer to Table 8) in ML tools when voltage was also considered a feature in GBM (see Fig. 11).



(a)



(b)

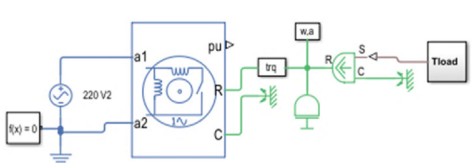
Fig. 7. Arc fault in the motor: (a) simulation model, (b) the effect on the current and THD.

Table 4. Machine learning result (Arc fault in Induction motor) by two features.

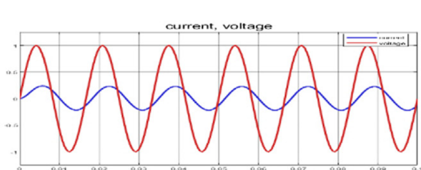
Accuracy	Precision	Recall	F1 score
98.39 %	99.34 %	98.24 %	98.79 %

2.3. Detecting faults in one system (Heating system)

Detecting three faults in a resistive load over different time intervals presents challenges, as each fault affects the voltage and current characteristics differently. Additionally, the THD varies according



(a)



(b)

Fig. 6. Single-phase induction motor: (a) block diagram, (b) voltage—current sine wave curve.

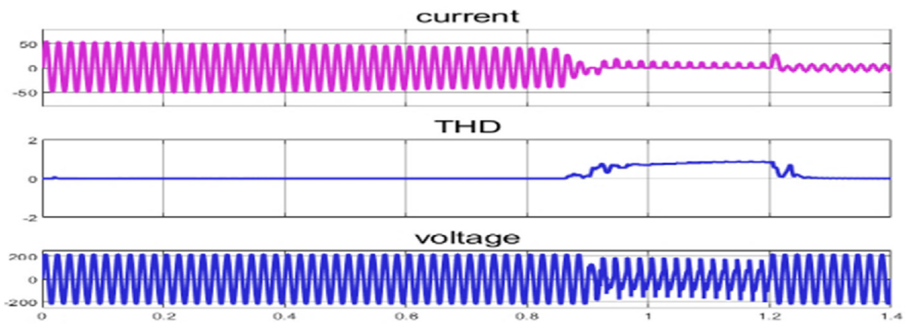


Fig. 8. The effect of overvoltage on the current, THD, and voltage on the motor.

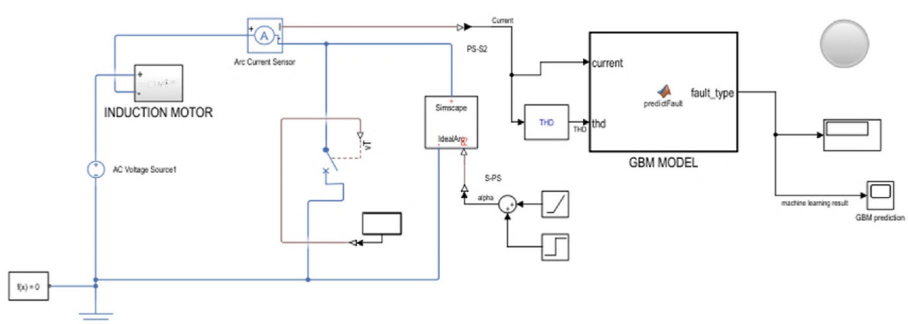
Table 5. Machine learning result (Arc fault in Induction motor) by three features.

Accuracy	Precision	Recall	F1 score
98.37 %	99.90 %	96.82 %	98.34 %

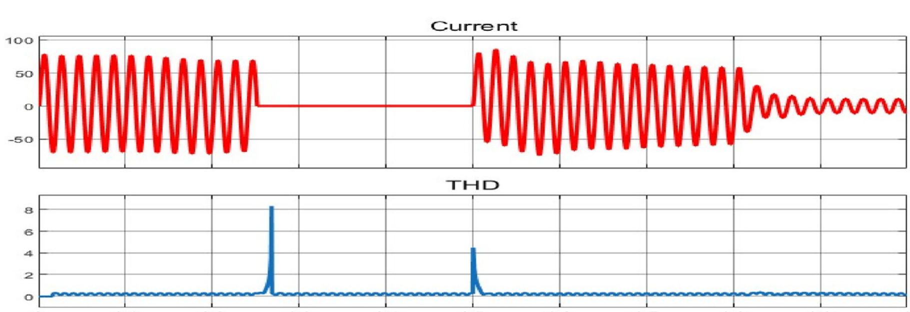
Table 6. Machine learning result (Contacting open fault in induction motor).

Accuracy	Precision	Recall	F1 score
98.65 %	98.24 %	99.08 %	98.66 %

to the type of fault. This is illustrated in Fig. 12, which shows three different types of faults occurring at different times. During the first interval, from 0 to 0.25 s, the system operates normally, exhibiting a standard sine wave curve for voltage and current. From 0.25 to 0.5 s, arc faults occur, making a shoulder for both voltage and current while reducing their amplitudes. Between 0.5 and 0.75 s, open contact faults result in no voltage or current across the resistive load. Finally, from 0.75 s to 1 min, overvoltage situations arise, significantly



(a)



(b)

Fig. 9. MATLAB Simulation for open contact fault in motor: (a) circuit diagram, (b) the effect of open contact fault on the current and THD in the induction motor.

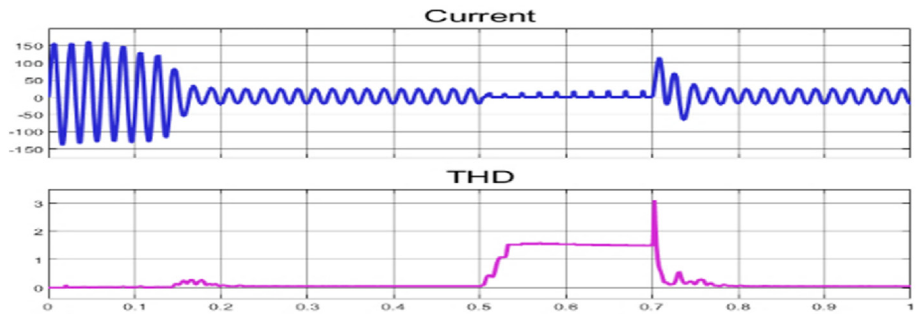


Fig. 10. Starting fault from 0.5 s to 0.7 s, and the effect on current and THD features.

Table 7. Machine learning result (Overvoltage fault in induction motor) by two features.

Accuracy	Precision	Recall	F1 score
96.15 %	94.93 %	97.70 %	96.30 %

Table 8. Machine learning result (Overvoltage fault in induction motor) by three features.

Accuracy	Precision	Recall	F1 score
96.99 %	94.89 %	99.31 %	97.05 %

impacting the THD. The accuracy of the GBM algorithm is 96.92 %, as shown in Table 9. This indicates that good accuracy in detecting each type of fault within the same heating system can be achieved using current and THD features.

Table 9. Machine learning result (Three faults in one system).

Accuracy	Precision	Recall	F1 score
96.92 %	94.28 %	99.92 %	97.02 %

2.4. Detecting arc fault in heater and induction motor

Selecting features for the simultaneous and parallel operation of two loads requires high accuracy due to varying current, voltage, and THD characteristics in different loads. The result of changing features in the MATLAB Simulink is shown in Fig. 13. It appears that there is an arc fault in the resistance from 0.1 to 0.25 s and an arc fault in the motor from 0.5 to 0.65 s. When the detection of arc faults focuses on four features (current on resistive

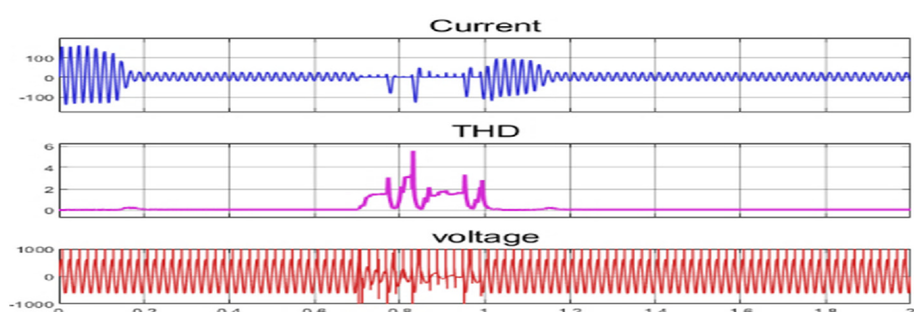


Fig. 11. Starting fault from 0.7 s to 1 s, the effect on current, THD, and voltage features.

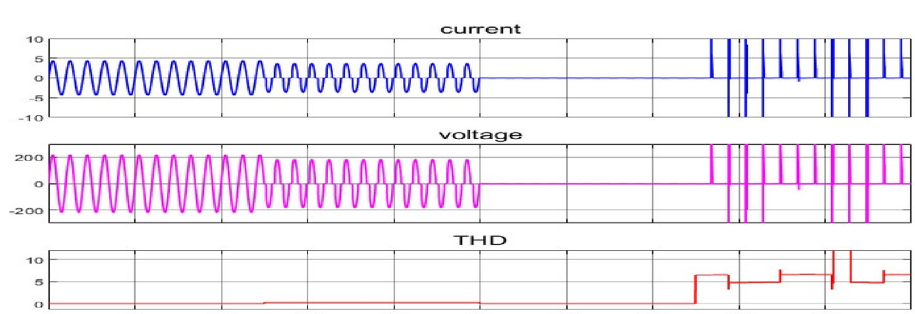


Fig. 12. Three faults on different periods (in 1 s) in current, voltage, and THD.

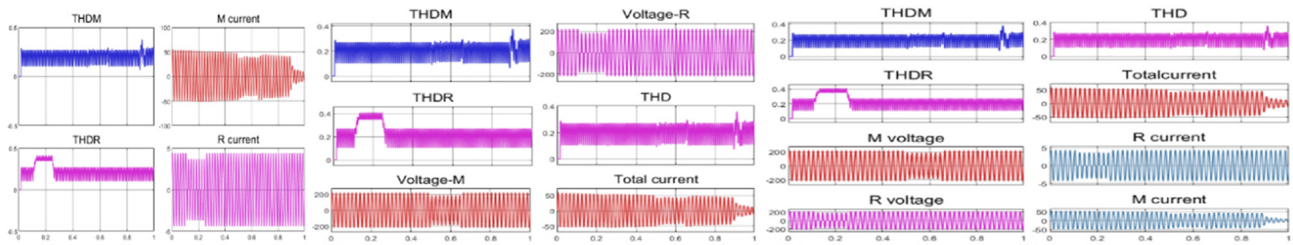


Fig. 13. Detecting arc fault on resistance and motor connecting in parallel (four features, six features, eight features).

load, the current in the induction motor, and THD in resistor and motor), the GBM algorithm hits 74.95 % accuracy (see Table 10). For checking and improving the accuracy, the algorithm performs six features (THD on the resistor and motor separately, the voltage on the motor and resistor separately, THD, and total current). This makes up for the fair detection by 77.84 % of the GBM algorithm. It revealed that the GBM's detection process is more accurate than the four features. In eight feature detections (THD in resistor and motor separately, voltage in motor and resistor separately, THD, total current, current in motor and resistive separately), the accuracy of the algorithm received 78.42 %.

2.5. Evaluation of gradient boosting machine

To evaluate the process, it is essential to utilize the same dataset generated from the heating system while applying the arc fault analysis. The data set has two features (current and THD) and one label (fault 0–1) and is moderate in size (500,002 rows).

Table 10. Machine learning result of Arc Fault in Heater and Induction Motor.

Metric	Four Features	Six Features	Eight Features
Accuracy	74.65 %	77.84 %	78.42 %
Precision	100.00 %	100.00 %	100.00 %
Recall	49.24 %	55.53 %	56.64 %
F1 score	65.99 %	71.41 %	72.32 %

2.5.1. Evaluation compared to other algorithms

These algorithms assess the efficiency and accuracy of predictive models across various domains. Gradient Boosting Machine (GBM) and Extreme Gradient Boosting (XGBoost) are effective ensemble methods for structured data, providing strong performance and good interpretability. Convolutional Neural Networks CNNs are primarily used for visual data, excelling in image classification, object detection, and facial recognition, but can also transform time-series or text data [15]. Support Vector Machines (SVMs) are versatile and used for classification, regression, and anomaly detection [16]. Long Short-Term Memory (LSTM) networks are ideal for sequential data, such as time-series analysis, speech recognition, and machine translation [17] (see Table 11).

2.5.2. Evaluation of time inference

Making a deep learning model lightweight reduces training time, storage, and high computational costs and ensures privacy preservation by eliminating the need for data transfer to the cloud [18]. Table 12 presents the configuration model and parameters for each algorithm, while Table 13

Table 11. Evaluation compared to other algorithms.

Algorithm	Hyperparameters	Accuracy	Precision	Recall	F1 Score
GBM	Number of trees, learning rate, max depth, subsampling.	98.15 %	98.42 %	97.85 %	98.14 %
XGBoost	Number of trees, learning rate, max depth, regularization, subsampling, feature sampling.	98.21 %	98.48 %	97.93 %	98.20 %
CNN	Number of filters, filter size, activation function, pooling size, learning rate, epochs, and batch size.	84.96 %	69.91 %	100 %	82.29 %
SVM	Kernel type, regularization parameter, gamma, degree (for polynomial kernel).	96 %	100 %	92.59 %	96.15 %
LSTM	MaxEpochs, MiniBatch, GradientThreshold, initial LearnRate, learn rate schedule, and DropFactor.	96.62 %	94.52 %	99 %	96.71 %

Table 12. Model Configurations for GBM and light model.

Model	GBM	Light_Model_Time
Tool	MATLAB's fitcensemble function	Decision tree classifier (fitctree in MATLAB)
Configuration	Used bagging with decision tree-based learners (MaxNumSplits = 10).	A single decision tree was selected as the lightweight model

includes the inference times. GBMs are noted for their high accuracy but tend to have longer inference times. In contrast, TinyML-based models focus on achieving faster inference and reducing computational overhead. It is essential to compare inference times across different feature sets for optimal model selection by the same data set.

3. Results and discussion

According to the outcomes of the proposed technique, the GBM algorithm effectively detects faults in electrical systems, identifying arcs, overvoltage, and open contact faults in both resistive and induction motor loads. Its high accuracy rates make it essential for fire detection in residential buildings. During an arc fault in a heating system, the current experiences fluctuations and decreases in amplitude. These changes in the current lead to significant variations in THD, which in turn affects the characteristics used for feature detection. Consequently, this enhances the accuracy of the GBM algorithm, resulting in an accuracy rate of 98.15 %. In the case of an open contact fault, the fault makes the current drop to nearly zero. This change influences the THD, creating two distinct pulses: one at the beginning and one at the end of the fault. The GBM can detect this fault with an accuracy rate of 98.81 %. For an overvoltage fault, the impact on the current is not prominently reflected in the sinewave curve. However, when examining THD, the detection process becomes more obvious, leading to an impressive accuracy of 99.53 % for the GBM. In an induction motor, the variation in current generates some harmonics, which are normal for the operating condition of the motor. Although these harmonics can create some confusion for the features of the algorithm, it still manages to detect the arc fault with an accuracy rate of 98.37 % and open contact faults with an accuracy of 98.65 %. For overvoltage faults, the algorithm checks two to three features, resulting in an improved performance from 96.24 % to 96.99 %. Detecting three types of faults in the heating system at different periods using only two features (current and THD) yields an impressive overall accuracy rate of 96.92 %. Two loads that connected parallelly (resistive and induction motor) had some challenges that affected the accuracy of detection of the faults, which resulted in to drop in the accuracy to just 74.95 % with four features, 77.85 % with six features, and 78.42 % with eight features. Effective feature selection played a crucial role in the performance of the GBM. Current, THD and voltage were utilized as features, demonstrating excellent capacity for fault detection. In contrast, the motor's fluctuating characteristics during startup required additional features, including voltage, to improve prediction accuracy. The investigation revealed that incorporating THD alongside either voltage or current led to a more precise classification of faults in varied operation scenarios (see Table 14).

Table 13. Time inference result.

Feature (set)	GBM_Time (sec)	Light_Model_Time (sec)
Current	5.9843	0.20078
THD	5.1499	0.20093
Current + THD	6.0441	0.059933

In GBM, feature importance is based on how much each feature reduces impurity at decision tree splits. These reductions are aggregated across all trees to determine the feature's overall importance, as shown in Fig. 14, where THD with larger reductions indicates greater significance.

Table 14. Summary table.

Fault Type	Load Type	Accuracy	F1 Score	Key trends
Arc fault	Heating system	98.15 %	98.14 %	The shoulder appears on the current sine wave
Arc fault	Induction motor	98.39 %	98.79 %	Decreasing amplitude of current
Contacting open	Heating system	98.81 %	98.81 %	Fault, starting with one impressive pulse and ending with the same pulse
Contacting open	Induction motor	98.65 %	98.66 %	The current crossing on the load is almost zero
Overvoltage fault	Heating system	99.53 %	99.53 %	Huge effect on the THD due to the increase and decrease in a short time
Overvoltage fault	Induction motor	96.24 %	96.30 %	Huge effect on the THD

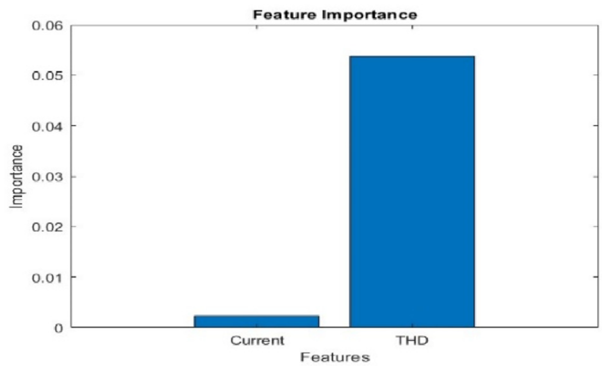


Fig. 14. Bar chart of feature importance.

Synthetic data has limitations, such as a lack of realism, biases, and validation issues. Future work should involve collecting real-world data, developing a hybrid testing framework, conducting on-site experiments, continuously refining models, and integrating safety monitoring. Collaboration with industry stakeholders is essential for gathering extensive datasets, creating a framework, performing field experiments, implementing adaptive mechanisms, and integrating with live monitoring systems for ongoing improvement.

GBM excels at efficiently handling structured or tabular data with minimal preprocessing, such as feature scaling and missing value management. It outperforms SVM with large datasets and captures complex patterns without kernel selection. GBM is faster to train than CNN and is ideal for non-spatial data applications. Although XGBoost provides enhancements like parallel tree construction and regularization, GBM remains simpler to implement and understand. Its boosting mechanism offers adaptability and strong performance in regression, classification, and ranking tasks. To validate GBM's effectiveness, it is essential to compare it with algorithms like XGBoost and Long Short-Term Memory (LSTM) networks, focusing on aspects such as computational efficiency, interpretability, and robustness.

Overall, the results show that the GBM algorithm, with its ability for high precision and adaptability through feature selection, provides a viable solution for real-time fault detection in electrical systems, thereby making a significant contribution to fire safety in residential buildings.

4. Conclusion

Early detection of fire significantly enhances safety in residential buildings. This paper presented the use of the Gradient Boosting Machine (GBM) algorithm for detecting electrical fires, specifically focusing on three different causes of these fire cases. The analysis demonstrates a substantial improvement in response times by analyzing voltage, current, and THD. The GBM algorithm notably enhances the accuracy of predictive analytics, particularly in the field of electrical safety. By analyzing historical data and identifying patterns, the algorithm can effectively detect potential electrical fire hazards before they occur. This proactive approach not only helps mitigate the risk of fires but also plays a vital role in reducing associated losses and damage to electrical circuits. Detecting electrical fires in specific loads becomes much more effective when evaluating each load individually. This targeted approach improves safety and helps identify potential hazards more quickly. However, the accuracy of the detection algorithm can decrease when identifying multiple loads with similar characteristics. This challenge can be addressed by enhancing the algorithm's features and improving its ability to recognize the distinct characteristics of each load, especially for Complex (or combined) Loads, which are a combination of resistive, inductive, and capacitive loads.

Author contributions

All authors contributed to the study's conception and design. MATLAB preparation, data collection, and analysis were performed by [S. A.], editing and supervision were performed by [F. A.]. The first draft of the manuscript was written by [S. A.].

Ethical statement

Not required.

AI usage declaration

None.

Funding information

No specific funding was received for conducting this research.

Conflict of interest

The authors declare that there is no conflict of interest.

Acknowledgment

Many thanks to Salahaddin University's Engineering College-Electrical Department for their helpful assistance. I would like to sincerely thank [Salahaddin University] for their kind financial assistance, which has been crucial to my research paper's successful completion.

References

[1] Yang Kai, Z R, Yang Jianhong, Liu Canhua, Chen Shouhong, Zhang F. A novel Arc fault detector for early detection of electrical fires. Sensors 2016 .2016;16:1−24. https://doi.org/10.3390/s16040500.

[2] Tian Chunpeng, LukunWang ZX, Liu Yunjie. Arc fault detection using artificial intelligence: challenges and benefits. Math Biosci Eng 2023;20(7):12404−32. https://doi.org/10.3934/mbe.2023552.

[3] Aziz Norshakirah, A EAP, Aziz Izzatdin Abdul. A study on gradient boosting algorithms for development of AI monitoring and prediction systems. In: 2020 international conference on computational intelligence (ICCI). IEEE Xplore; 2020. p. 11−6. https://doi.org/10.1109/ICCI51257.2020.9247843.

[4] Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. Artif Intell Rev 2021;54:1937−67. https://doi.org/10.1007/s10462-020-09896-5.

[5] Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorob 2013;7(21). https://doi.org/10.3389/fnbot.2013.00021.

[6] Markoulidakis I, Markoulidakis G. Probabilistic confusion matrix: a novel method for machine learning algorithm generalized performance analysis. Technologies 2024 .2024; 12:1−23. https://doi.org/10.3390/technologies12070113.

[7] Andrea Jonathan, B P, Zirn Olivier, Bournat Marc. The Electric Arc as a Circuit Component. IEEE; 2015. p. 3027−34. https://doi.org/10.1109/IECON.2015.7392564.

[8] Han Xiangyu, L D, Huang Lizong, Huang Hanqing, Yang Jin, Zhang Yilei, et al. Series Arc fault detection method based on category recognition and artificial neural network. Electronics 2020;9:1367. https://doi.org/10.3390/electronics9091367.

[9] Gao Hongxin, W Z, Tang Aixia, Han Congxin, Guo Fengyi, Li Baifu. Research on series Arc Fault detection and phase selection feature extraction method. IEEE ASME Trans Mechatron 2021;70. https://doi.org/10.1109/tim.2021.3080376. 2004508-2004508.

[10] Jiang Jun, M S, Li Wei, Wen Zhe, Bie Yifan, Schwarz Harald, et al. Series Arc fault detection based on random forest and deep neural network. IEEE Sensor Council 2021. https://doi.org/10.1109/jsen.2021.3082294.

[11] Han Congxin, W Z, Tang Aixia, Gao Hongxin, Guo Fengyi. Recognition method of AC series Arc Fault characteristics under complicated harmonic conditions. IEEE Trans Instrum Meas 2021;70. https://doi.org/10.1109/tim.2021.3051669. 3509709-3509709.

[12] Wang H, Kang J, Lin Y. Low-voltage series Arc Fault detection based on multi-feature fusion and improved residual network. Electronics 2025;14(7):1325. https://doi.org/10.3390/electronics14071325.

[13] Lin XJ, Z N, Mao YH, Chen b JY, Tian XT, Zhong W. A review of the transformation from urban centralized heating system to integrated energy system in smart city. Science (New York, N Y) 2024;240:122272. https://doi.org/10.1016/j.applthermaleng.2023.122272.

[14] Ray MP. Electrical devices & systems. In: Sreekanth Reddy K, editor. CIIR research publications; 2023. p. 41−4 [INDIA: UTTAR PRADESH].

[15] Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 2021;8:53. https://doi.org/10.1186/s40537-021-00444-8.

[16] Awad M, Khanna R. Support vector machines for classification. In: Efficient learning machines; 2015. https://doi.org/10.1007/978-1-4302-5990-9_3.

[17] Van Houdt G, Mosquera C, Nápoles G. A review on the long short-term memory model. Artif Intell Rev 2020;53:5929−55. https://doi.org/10.1007/s10462-020-09838-1.

[18] Dutta S, Das M. Maulik. Lightweight deep learning models for aerial scene classification: a comprehensive survey. Eng Appl Artif Intell 2025;142:109859. https://doi.org/10.1016/j.engappai.2024.109859.