Early outlier detection in three-phase induction heating systems using clustering algorithms

Mohammed H. Qais a, *, Seema Kewat b, K.H. Loo c, Cheung-Ming Lai a

a Center for Advances in Reliability and Safety, Hong Kong, China
b Department of Engineering and Physics, Karlstad University, Karlstad, Sweden
c Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Keywords:
Outlier detection
Clustering algorithms
Induction Heating
K-means
Unsupervised machine learning

ABSTRACT

Induction heating (IH) devices transfer the electric power to the contactless cookware via the electromagnetic field. Therefore, the temperature of cookware is measured remotely, and the early detection of cookware overheating will ensure the user’s safety as well as extend the remaining useful life of electronic components. In this work, we present a clustering model for outlier detection in IH systems based on clustering algorithms and measured data using two thermal sensors. First, a healthy dataset is collected for the temperatures of inverters and cookware under different sizes and materials of cookware items, different amounts of water in cookware, and different amounts of electrical power. After that, K-means and fuzzy c-means were utilized to cluster this normal dataset, where the maximum distance between their centers and data points was selected as a threshold. Finally, the clustered model is evaluated using a testing dataset that includes outliers. According to the results, the K-means algorithm detected around 96% of the produced outliers, however, the fuzzy c-means algorithm detected around 68%. In conclusion, the deployment of the clustering model in outlier detection is simple and uses only the threshold and the cluster centers.

1. Introduction

In recent years, the technology of induction heating systems has improved and prices have decreased, attracting more chefs and residential cooks [1]. Even though it is expensive at first, its high efficiency (94% [2]) compared to gas stoves (50% [3]) can make up for this. This is because the cookware itself is the source of heat on an induction cooktop [4]. In addition, induction cooktops have several advantages, including no harmful CO2 emissions [5], safe contactless heating, smaller dimensions, and an extremely fast heating time [6]. The induction heating system mostly uses an AC/DC/AC converter to convert the low-frequency (50/60 Hz) AC input into the higher-frequency (20—100 kHz) AC output [7]. This converter includes power electronics devices, such as insulated-gate bipolar transistors (IGBTs), that are precisely controlled to deliver a specific amount of power to an induction coil [8]. The power will then be transferred to the cookware via an electromagnetic field [9], which produces eddy currents on the bottom surface of the cookware [10]. Finally, heat energy is generated in the cookware by these circulating currents.

1.1. Research motivation

Cookware’s temperature rise depends on several variables, including the cookware’s material, size, placement on the top of the induction coil [11], amount of power given, and the amount of food or water being cooked [12]. However, the behavior of end-users varies and cannot be predicted. Therefore, manufacturers try to strengthen their products by increasing the safety margin of the system rating, which will increase the cost. The induction cooktop industry is rapidly evolving and becoming more intelligent, particularly in terms of user safety. However, in the literature, the researchers mostly focused on the design of inverter topologies [13], such as single-switch inverters, half-bridge inverters, and full-bridge inverters, according to the power rating [14]. In addition, they focused on the control system of inverters that is responsible for regulating the amount of output power as well as protecting the induction cooktop during faulty conditions [15]. Moreover, they presented advanced recognition techniques for the size and material of cookware [16]. However, there is a research gap in outlier detection in induction heating systems. The manufacturers still rely on conventional techniques [17], such as using a preset threshold value. However, few
advanced artificial intelligence-based early outlier detection methods are used in induction heating systems [18].

1.2. Literature review

In the literature, various methods for outlier detection were used, including statistical-based methods [19] and unsupervised machine learning methods [20,21]. The accuracy of these methods varies depending on the amount of data, the number of features, and the type of application (e.g., time series or images). Statistical methods include mean and standard deviations, the Gaussian model-based method [22], and the Mahalanobis distance [23]. Clustering algorithms [24], deep learning-based auto-encoders [25], and one-class classifiers are examples of unsupervised machine learning [26]. Clustering algorithms are used only for two-dimensional features. As a result, other methods, for example, principal component analysis (PCA) [27], can be used alongside clustering algorithms to reduce the higher dimensional features. However, this reduction can eliminate or hide some important features, which can reduce detection accuracy. Because only two features (two thermocouple sensors) will be used in this work, clustering algorithms can be applied smoothly for outlier detection in induction heating systems.

Due to their ease of deployment, unsupervised clustering algorithms have been widely applied to find outliers in numerous engineering disciplines [28], such as outliers in district heating substations [29] and outliers in building energy consumption [30]. These methods cluster the collected data points from only two sensors. The precision of these methods increases with the number of datasets. The clustering-based outlier detection methods first cluster normal data collected under various operating settings to determine the maximum distance, or threshold, between data points and their centroids. After that, any measured point with a distance greater than this threshold will be identified as an outlier. Many clustering methods have been applied to find outliers, such as the density-based spatial clustering algorithm [31], the K-means clustering algorithm [32], and the fuzzy c-means clustering algorithm [33].

1.3. Main contribution

According to the knowledge of the authors, this is the first time an outlier detection method based on a clustering algorithm has been implemented in an induction heating system. This article uses K-means and fuzzy c-means clustering algorithms to classify the massive quantity of data collected from two temperature sensors.

The main contributions of this paper are summarized as follows:

1. Proposing an unsupervised learning model to detect outliers and enhance the users’ security based on the measured data using two thermal sensors.
2. Utilizing the clustering algorithms, K-means and fuzzy c-means, to find the centers of the clustered healthy data and find the maximum distance between centers and healthy data points.
3. Applying and testing the proposed model to find outliers during abnormal operating conditions.

The rest of the paper is organized as follows: Section 2 describes the modeling of the induction heating system. The clustering algorithms K-means and fuzzy c-means are discussed in Section 3. Section 4 presents details of the components of a commercial induction cooktop. Section 5 describes the proposed outlier detection method. Section 6 shows the operational graphs of the three-phase induction heating system.
results and discussions.

2. Three-phase induction heating system

As shown in Fig. 1(a), the induction heating (IH) device converts the low-frequency AC power source (50/60 Hz) into a high-frequency AC power output (>20 kHz). A high-power IH device gets its electricity from a three-phase AC source. The AC power is converted into DC power by a rectifier. A DC capacitor (C_Dc) is used to smooth out the rectifier’s output DC voltage. The DC voltage (V_Dc) is then converted to high-frequency AC power (V_o) with the help of the resonant inverter. This resonant inverter includes two insulated-gate bipolar transistors (IGBTs) (Q1 and Q2) that are connected with opposite diodes and a resonance capacitor (C_r). The amount of output AC power can be controlled by the switching frequency of the IGBTs using pulse width modulation (PWM) signals (g1 and g2). Then, the output power is transmitted wirelessly to the cookware through the electromagnetic field that is produced by the cookware through the electromagnetic field that is produced by the alternating magnetic field intersects it. Then, the eddy currents (i_eq) will be produced on the bottom surface of the cookware when the alternating magnetic field intersects with it. Then, the eddy currents (i_eq) will be produced on the bottom surface due to the closed resistance (R_eq) on it. Finally, according to Joule’s law, heat energy is generated on the bottom surface of the cookware. However, the output heat energy cannot be measured directly in the IH device due to the contactless operation of cookware, as shown in Fig. 1(b). Therefore, an equivalent circuit of coil cookware, including equivalent resistance and inductance (R_eq and L_eq), is constructed as shown in Fig. 1(c). Moreover, the equivalent circuit of the inverter is a dependent voltage source of collector-emitter voltage (V_CE) and on-state resistance (R_eq).

2.1. Cookware temperature

In IH systems, the energy conversion from electricity to heat occurs on the cookware itself, so the efficiency is very high. However, there are many factors in the cookware material and sizes that affect the amount of produced heat energy. For induction cookware, ferromagnetic materials, such as stainless steel and cast iron, are used because of their high relative magnetic permeability (μ_r). The value of the eddy current (i_eq) decreases with the increase of the skin depth (δ) of the bottom surface of the cookware, as seen in (1). Moreover, the frequency (f) of the eddy current has an inverse relation with the skin depth, as seen in (2). Consequently, the value of skin depth has an inverse relationship with the surface resistance (R_eq) of cookware, as seen in (3). Moreover, the resistivity (ρ) of cookware material has an impact on its surface resistance. As a result, the flow of eddy currents through the surface resistance of cookware causes power losses as a form of heat energy (Q_eq), as seen in (4). The heat energy increases with the decrease of skin depth due to the increase in frequency.

\[ i_{eq} = i_{(max)}e^{-\frac{x}{\delta}} \]  
\[ \delta = \sqrt{\frac{1}{4\pi \times 10^{-7}}} \sqrt{\frac{\rho}{\mu_r}} \]  
\[ R_eq = \frac{\rho}{\delta} \]  
\[ Q_eq = \dot{i}_eq(t)R_eq = \dot{I}_eq(t)R_eq \] (4)

where \( \omega \) is the angular frequency, \( R_1 \) is the coil resistance, \( L_2 \) is the coil inductance, \( L_2 \) is the cookware inductance, and \( M \) is the mutual inductance. Therefore, the total output power can be calculated using the flow of output current (i_eq) through the equivalent resistance, as seen in (6), which is approximately equivalent to the power loss due to the flow of eddy currents through the surface resistance of cookware. Then, the cookware temperature (T_k) can be estimated from (7) [34], which is affected by the frequency, the cookware material and size, the output current, and the equivalent resistance.

\[ \dot{e}_eq(t)R_eq \approx \dot{i}_eq(t)R_eq \] (6)
\[ \Delta T_k = \frac{1}{mC_p} \int_0^t \dot{e}_eq(t)R_eq \, dt \] (7)

2.2. Inverter module temperature

As discussed above, the inverter module is made of two IGBTs and two diodes. The switching and conduction of IGBTs cause power losses (P_D) as a form of heat energy that increases the junction temperature of IGBTs (T_J). The total power losses (P_T), as seen in (8), can be approximated to P_D because it is very high compared to the diode power losses (P_D) [35].

\[ P_T = P_D + P_D \approx P_D \] (8)
\[ P_D = \frac{1}{T} \int_0^T v_{CE}(t)i_c(t)D(t)dt \] (9)
\[ \Delta T_k = \int_0^T \frac{\partial Z_{th}(t)}{\partial t}dt \] (10)

where D is the duty cycle, \( v_{CE} \) is the collector-emitter voltage, \( i_c \) is the collector current, and \( Z_{th} \) is the thermal impedance between the junction of the IGBT and the heat sink. During the on-state of IGBT, the collector current is the same value as the output current (i_eq).

In conclusion, from (6) and (9), both \( T_F \) and \( T_G \) are relying on the amount of output current, so there is a correlation between these temperatures that can be utilized in this work. However, we cannot find a mathematical expression that correlates both temperatures because there is no fixed cookware and no fixed liquid level inside it. Therefore, the learning algorithms are the best choice to predict these temperatures based on learned models using large training data for \( T_F \) and \( T_G \).

3. Clustering algorithms

Clustering is used to divide a set of N data points \( X = \{x_1, x_2, \ldots, x_N\} \in \mathbb{R}^d \) into M clusters’ centroids \( C = \{c_1, c_2, \ldots, c_M\} \in \mathbb{R}^d \) that are compact and well separated from one another. A common method for assessing the quality of cluster analysis is to minimize the sum of squared Euclidean distances (D) between the data points and their respective centroids (c).

\[ D(x, c_i) = ||x_i - c_i||^2 \] (11)

There are two common clustering types: hard and soft clustering methods. Hard clustering means that every data point belongs to only one cluster, such as the K-means algorithm [36]. However, soft clustering means that every data point is part of all clusters but with different membership degrees, such as in the Fuzzy c-means algorithm [37]. Therefore, in this work, we applied both types of clustering algorithms for comparison.

3.1. K-Means

The K-means approach is widely used in many clustering applications because it is a very fast iterative method and has a simple structure. The K-means algorithm starts with initial centroids of k-clusters that are
selected randomly from the data points. Then, K-means divides the data points into $k$ clusters by minimizing the loss function $J$ in (12).

$$J = \min \sum_{i=1}^{N} \sum_{k=1}^{M} I_k D_k(x_i, c_k)$$  \hspace{1cm} (12)

$$I_k = \begin{cases} 1 & \text{if } \|x_i - c_k\| = \min_{1\leq k'\leq M} \|x_i - c_{k'}\| \\ 0 & \text{otherwise} \end{cases}$$  \hspace{1cm} (13)

After that, K-means uses the Lloyd algorithm to update the center $c_k$ of $k$-clusters by finding the average value of all data points in the same cluster, as in (14).

$$c_k = \frac{\sum_{i=1}^{N} I_k x_i}{\sum_{i=1}^{N} I_k}$$  \hspace{1cm} (14)

The main challenge of K-means, however, is the centroid initialization; thus, K-means++ is proposed. K-means++ first selects centroids at random and then computes the Euclidean distance $D_k(x_i, c_k)$ of all data points, where the minimum distance indicates that $x_i$ belongs to a $k$-cluster. After that, all centroids will be updated according to the maximum probability $P(x_i)$ of clustered data points ($x_i$) as follows:

$$p(x_i) = \frac{D_k(x_i, c_k)}{\sum_{j=1}^{M} D_k(x_i, c_j)}$$  \hspace{1cm} (15)

The steps of the K-means algorithm are described by using the pseudocode, as shown in Algorithm 1.

**Algorithm 1**  k-means algorithm

1. Input the dataset $\chi$ and the number of $k$-clusters.
2. Initialize the centroids using the K-means++ method.
3. Find the loss functions as in (12).
4. Update the centers of clusters using (14).
5. Repeat steps 4 and 5 until the $\Delta J < \epsilon$ or reaches the number of maximum iterations.
6. Return the centers and clustered data.

### 3.2. Fuzzy c-means

This clustering algorithm is considered a soft method, where all data points belong to all clusters with different membership degrees $\mu$. Therefore, the objective function of fuzzy c-means (FCM) will include $\mu$ as follows:

$$J = \min \sum_{i=1}^{N} \sum_{k=1}^{M} \mu_{ik}^m D_k(x_i, c_k)$$  \hspace{1cm} (16)

$$\mu_{ik} \in [0,1], \sum_{k=1}^{M} \mu_{ik} = 1, \text{ and } 0 < \sum_{i=1}^{N} \mu_{ij} < N$$  \hspace{1cm} (17)

where $m$ is the fuzzification coefficient (where $m > 1$), The FCM will minimize the loss function by updating $c$ and $\mu$ as in (18) and (19). It will continue till they reach the maximum iterations.

$$c_k = \frac{\sum_{i=1}^{N} (\mu_{ik})^m x_i}{\sum_{i=1}^{N} (\mu_{ik})^m}$$  \hspace{1cm} (18)

$$\mu_{ik} = \left( \frac{D_k(x_i, c_k)}{\sum_{j=1}^{M} D_k(x_i, c_j)} \right)^{\frac{1}{m-1}}$$  \hspace{1cm} (19)

The pseudocode of the FCM algorithm is shown in Algorithm 2.

**Algorithm 2**  Fuzzy c-means algorithm

1. Input dataset $\chi$, $k$-clusters, and $m$.
2. Random initialization of $M = [\mu_{ik}]_{N \times M}$, $\mu_{ik} \in [0,1]$.
3. Find the center matrix $C$ by Eq. (18).
4. Update the membership matrix $\mu$ by Eq. (19).
5. Repeat steps 3 and 4 until reaching maximum iterations or the tolerance error.
6. Return the centroids and membership matrices $C$ and $\mu$.

### 4. Early outlier detection methodology

Many variables can cause failure in induction heating systems, such as the environment, bad behavior by chefs, unsuitable cookware, and degradation of the system’s components. Mostly, the manufacturers are worried about the undesired operating conditions; therefore, they try to use higher power-rated components and install very fast protective components, such as fuses, that require maintenance engineers, adding cost. Recently, artificial intelligence (AI) methods can provide early fault detection and prevent the system from failure, as well as assure the users’ safety. These AI-based outlier detection methods have mostly been trained previously on large amounts of normal data. Then, these methods will detect the abnormal condition in its early stage, which will turn off the system and send an alarm to the interface panel. There are two important NTC sensors installed in induction heating systems to measure the pan temperature and the case temperature of the IGBT module. These temperatures vary depending on the power level, the size and material of the cookware, the position of the cookware on the induction coil, and the amount of water or food in the cookware.

#### 4.1. Methodology

In this work, we applied the following methodology, as shown in Fig. 2, to propose an early outlier detection method, as follows: first collecting a large dataset of temperatures under various normal operating conditions; then, applying clustering algorithms, K-means, and FCM, to cluster the collected dataset into different clusters that have similar features; after that, we can find the threshold value by finding the maximum Euclidean distance ($D_k$) between data points and their centroids; finally, we proposed an outlier detection algorithm for induction heating systems.

#### 4.2. Deployment method

After clustering a large amount of collected healthy data into $k$-clusters and finding the threshold value, we will deploy this AI model in the commercial induction heating system. The measured temperatures of $T_C$ and $T_P$ by NTC sensors placed in the heat sink of the IGBT and under the bottom surface of the cookware, respectively, are the inputs of this AI model. First, the input temperatures $T_C$ and $T_P$ will be divided by preset temperature values. In this work, they are 50 and 100 °C. After that, the Euclidean distance ($D_k$) between these normalized data points $x$ ($T_{C \circ P}$) and all centroids $c(T_{C \circ P})$ will be calculated as follows:

$$D_k = \sqrt{(x(T_C) - c_1(T_C))^2 + (x(T_P) - c_1(T_P))^2}$$  \hspace{1cm} (20)

Then, we will select the minimum distance from the $k$-distances between the datapoint and $k$-centroids to locate the measured point in the appropriate cluster. Then, this distance will be compared with the defined threshold value. If the distance is greater than the threshold, the model will identify it as an outlier and send an alarm to the interface panel, and at the same time, it will turn off the induction heating system for a while, as shown in Fig. 3.

#### 5. Experimental results and discussions

Recently, induction heating systems have been used widely in domestic, commercial, and industrial applications. The power circuit topology varies based on the power rating. In this study, a 15-kW high-power induction cooktop is investigated, as shown in Fig. 4(a). This system setup includes (A) a selector switch for power level selection ($P_s$); (B) a box of power and control boards; (C) an induction coil placed under a glass surface inside a big box; (D) Above the induction coil, cookware can be placed on top of the glass surface; (E) Thermocouples to measure the pan temperature ($T_P$) and case temperature ($T_C$) of the IGBT module;
Fig. 2. Flowchart methodology of proposing outlier detection method for induction heating system.

Fig. 3. Deployment of proposed outlier detection method for induction heating system.

and (F) Temperature acquisition device (SH-X).

The experimental setup of the power and control board is shown in Fig. 4(b), and its circuit diagram is shown in Fig. 4(c). The main input is a three-phase AC system, where the line-to-line voltage ($V_{L-L}$) is 380 V and the frequency is 50 Hz. The power circuit of the induction heating system includes: (I) A three-phase diode-bridge rectifier; (II) The DC link capacitor is connected to the output of the rectifier, where its capacitance value ($C_d$) is 30 μF and the DC voltage ($V_{DC}$) is approximately 500 V; (III) The filter capacitor is used, where its capacitance value ($C_f$) is 0.474 μF; (IV) The DC voltage is converted to high-frequency AC voltage using a half-bridge inverter circuit that consists of two IGBTs (1200 V and 200 A); (V) Snubber circuits ($C = 30 \, nF$ and $R = 300 \, kΩ$) are connected in parallel with the IGBTs; (VI) A high-voltage resonance capacitor, where its capacitances are given by ($C_{1} = C_{2} = 0.7 \, μF$); (VII) the Litz-Wire planar induction coil, which has an equivalent inductance of ($L_{eq} = 85 \, μH$) and resistance of ($R_{eq} = 0.3 \, Ω$), is connected to the output of resonant inverter; and (VIII) The main control unit (MCU) regulates the switching frequency (50–21 kHz) to change the output power (6.5–15 kW), as shown in Fig. 5. Moreover, MCU monitors the condition of the induction heating system using different sensors, including the DC current sensor ($I_{DC}$), the collector-emitter voltage ($V_{CE}$) sensor, the negative temperature coefficient (NTC) sensor to measure the bottom surface temperature of the pan or pot ($T_p$), and the case temperature of the IGBT module ($T_c$).

5.1. Healthy data collection

This study includes most of the normal operating conditions, cookware sizes and materials, different behaviors, and different water levels. The induction heating system has eight power levels (6.5–15 kW), as shown in Fig. 4. The select switch allows us to choose between these power levels. These levels can be used by end-users based on their needs and cooking requirements. To begin, this induction system is mostly used in restaurants to prepare large amounts of soup. As a result, we turned off the cooktop at power level 6 since the temperature rises extremely quickly with this water level (5%) and may reach an abnormal degree. The temperature rise of $T_c$ is shown in Fig. 7(b), where the rate of rise is different after fan operation at 50 °C. Additionally, we must gather more data in various sizes (50 × 50 and 30 × 30 cm). Fig. 8 displays the temperature responses of both cookware items at various power levels. As can be seen in Fig. 8(a), the pan temperature rises more quickly for smaller sizes. On the other hand, the $T_c$ rises more quickly for smaller cookware sizes than for bigger sizes, as displayed in Fig. 8(b). Finally, the material of cookware affects the cooktop functions, where only ferromagnetic materials like iron and stainless steel can be used. In this study, we investigated the temperature responses of two pieces of cookware of the same size (11 × 23 cm) but made of different materials: iron-cast and stainless steel. Fig. 8(a) and (b) demonstrate that the rising rates of $T_p$ and $T_c$ in an iron-cast pan are greater than those in a stainless-steel pan.

5.2. Data clustering

At this stage, all acquired $T_p$ and $T_c$ for all normal operating conditions are paired in a single two-column array dataset. The resulting size of the dataset is 2 × 120,000. This dataset is unlabeled and healthy because all data points were collected under all possible normal and healthy operating conditions. Following that, the data is normalized using the temperature presets for cooling fan operation. As a result, the $T_p$ column is divided by 100 °C, and the $T_c$ column is divided by 50 °C. The clustering algorithms K-means and FCM are then used to divide the data into clusters. The clustering algorithms of Scikit-learn are used, and the experiments are executed in Jupyter on a Windows 10 64-bit, Core 17 16 GB RAM laptop. Fig. 10 shows the sum of squared distances between data points and centers with the change in the number of clusters, where 14 are enough clusters to achieve good results, which is consistent with the various operating conditions. Table 1 shows the centers of the 14 clusters, where $T_{C(0)}$ and $T_{P(0)}$ are the normalized temperatures of $T_c$ and $T_p$, respectively. The positions of all clustered data points $x(T_{C(N)}, T_{P(N)})$ using K-means and FCM algorithms are shown in Fig. 11(a) and 11(b). K-means and FCM cluster the data by minimizing the values of objective functions (12) and (16), respectively. They first compute the Euclidean distances between a datapoint and all centers, and then assign this datapoint to a cluster if the distance to its center is the shortest. After that, the centers for K-means and FCM will be updated using (14) and (18), respectively. These steps will be repeated until the
Fig. 4. Experimental setup in Lab; (a) Induction Cooktop; (b) Main power circuit and control; and (c) Circuit Diagram of the main power circuit.
Fig. 5. Eight levels of the output power of a 15-kW commercial induction cooktop.

Fig. 6. Temperature rise for each power Level; (a) $T_p$; and (b) $T_C$.

Fig. 7. Temperature rise for different water Levels with different power levels; (a) $T_p$; and (b) $T_C$. 
5.3. Threshold identification

The Euclidean distances between all clustered healthy data points and their centroids are stored during the clustering stage. Fig. 12 shows the histogram distribution of all Euclidean distances. The average value of distances between data points and their centroids is 0.05. The maximum distance using K-means and FCM clustering are 0.219 and 0.2588, respectively, as shown in Fig. 12(a) and (b). These maximum values are defined as “threshold values”, where any distance greater than these values will be considered an outlier.

<table>
<thead>
<tr>
<th>Centroid No.</th>
<th>T_C(N)</th>
<th>T_P(N)</th>
<th>T_C(N)</th>
<th>T_P(N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.458119</td>
<td>0.283955</td>
<td>0.453347</td>
<td>0.262414</td>
</tr>
<tr>
<td>2</td>
<td>0.611995</td>
<td>0.57491</td>
<td>0.605825</td>
<td>0.602425</td>
</tr>
<tr>
<td>3</td>
<td>0.673839</td>
<td>0.714746</td>
<td>0.688141</td>
<td>0.723525</td>
</tr>
<tr>
<td>4</td>
<td>0.730713</td>
<td>0.864041</td>
<td>0.717037</td>
<td>0.575401</td>
</tr>
<tr>
<td>5</td>
<td>0.736698</td>
<td>0.584012</td>
<td>0.733591</td>
<td>0.891411</td>
</tr>
<tr>
<td>6</td>
<td>0.797825</td>
<td>0.969697</td>
<td>0.80628</td>
<td>0.649426</td>
</tr>
<tr>
<td>7</td>
<td>0.828085</td>
<td>0.656577</td>
<td>0.937904</td>
<td>0.737586</td>
</tr>
<tr>
<td>8</td>
<td>0.958864</td>
<td>0.758409</td>
<td>0.937583</td>
<td>0.900646</td>
</tr>
<tr>
<td>9</td>
<td>0.966405</td>
<td>0.867033</td>
<td>0.980177</td>
<td>1.105619</td>
</tr>
<tr>
<td>10</td>
<td>0.973024</td>
<td>1.119591</td>
<td>0.984677</td>
<td>0.815329</td>
</tr>
<tr>
<td>11</td>
<td>0.997611</td>
<td>1.004226</td>
<td>0.998886</td>
<td>1.006018</td>
</tr>
<tr>
<td>12</td>
<td>1.055235</td>
<td>1.248895</td>
<td>1.052886</td>
<td>1.253647</td>
</tr>
<tr>
<td>13</td>
<td>1.07294</td>
<td>1.128384</td>
<td>1.056561</td>
<td>1.132996</td>
</tr>
<tr>
<td>14</td>
<td>1.150853</td>
<td>1.33443</td>
<td>1.150779</td>
<td>1.333839</td>
</tr>
</tbody>
</table>

Table 1
The Centroids of 14-clusters.
5.4. Outlier detection verification

To investigate the robustness of the proposed outlier detection method, we collected a new dataset that includes both healthy and anomalous conditions. First, we presented some temporary, nondestructive outliers, where we switched off the induction heating system immediately after the intended abnormal temperature rise. These outliers include disconnecting the cooling fan of the IGBT, poor air ventilation in the box containing the power and control boards, and increasing the pan temperature out of the normal range by using small cookware at high power levels or using very little water at high power levels. Then, we applied the centroids of the 14 clusters in Table I to cluster the new dataset. Fig. 13(a) and 13(b) show the positions of 14 clusters using the centroids of K-means and FCM, respectively. As shown in Fig. 11.

Fig. 11. Healthy data clustering using; (a) K-means; and (b) FCM.
in Fig. 13, all Euclidean distances are compared to the threshold value, and data points with distances greater than the threshold are labeled as outliers in red.

The outliers in region A in Fig. 13 are due to the abnormal rising rate of pan temperature due to incompatible pan size and power level or a very small amount of water with a long heating time. However, the outliers in Region B are due to the failure of the cooling fan or very poor ventilation in the power and control box.

5.5. Discussion

Based on the temperature data collected in Section V(A), we found that the rise rates of temperatures in the pan and in the case of the IGBT can be paired to detect outliers. The time series of the normalized pan temperature (right brown axis) and the normalized IGBT case temperature (left blue axis) are shown in Fig. 14. If the rate of rise of the pan temperature differs from that of the case temperature, the proposed
Fig. 13. New data clustering using: (a) K-means; and (b) FCM.
method will identify it as an outlier, as shown in regions (A and B) of Fig. 14. Moreover, if the rate of rise of the case temperature differs from that of the pan temperature, it will be detected as an outlier, as shown in region (C) in Fig. 14. There is another region (D) where the case temperature is high, but it is identified as a normal operation when compared to the pan temperature.

The reason we chose 14 clusters for this work needs to be discussed in more detail. The answer is that we experimented with various cluster numbers and evaluated the accuracy of outlier detection for each algorithm and cluster size. The ratio of the number of detected anomalous data points to the actual number of anomalous points is the accuracy. Approximately 3200 actual anomalous data points were labeled in this study. So, to identify these anomalous data at various cluster sizes, we used K-means and FCM. The accuracy fluctuation with cluster counts for K-means and FCM is shown in Fig. 15. It is obvious that the maximum precision was achieved at 14 clusters. We also found that the accuracy was greater with K-means (96% vs. 68% for FCM). In actuality, 14 clusters make sense because we gathered data using various power levels (8 levels), materials, pan sizes, and water depths.

6. Conclusion

In this paper, we apply unsupervised machine-learning algorithms for early outlier detection in high-power induction heating systems.

---

Fig. 14. Detected abnormal temperature behavior using: (a) K-means; and (b) FCM.
First, two thermal sensors were utilized to collect large amounts of data on the temperatures of cookware and inverter modules at different healthy operating conditions. These healthy conditions include all power levels, cookware sizes and materials, and different water levels in cookware. After that, we applied clustering algorithms, K-means, and fuzzy c-means, to cluster the healthy data into many groups.

In this work, we found that 14 clusters are the best choice, which is consistent with the number of operating conditions. After data clustering, we determined the threshold value, which is the maximum Euclidean distance from data points to the centers of clusters, that can identify anomalous data points.

Finally, to investigate the robustness of the proposed outlier detection method, we collected new data that included anomalous and healthy data. The obtained results revealed that the proposed method based on the temperatures of cookware and the inverter can detect outliers early, before system failure. Moreover, the results show that the accuracy of the deployment of the proposed outlier detection based on the K-means algorithm is 96%. For future work, we will apply a transformer model for outlier detection and compare it with the clustering model.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is supported by Centre for Advances in Reliability and Safety (CAiRS) admitted under AIR@InnoHK Research Cluster.

References


Fig. 15. Outlier detection accuracy with cluster numbers.
14

M. H. Qais received his B.Sc. degree in Electrical Engineering from Sana’a University, Yemen, in 2007 and his M.Sc. and Ph. D. degrees in Electrical Engineering from King Saud University, Saudi Arabia, in 2014 and 2020, respectively. Since August 2021, he has been a postdoctoral fellow with the Centre for Advances in Reliability and Safety Limited (CAiRS), Hong Kong, China. From June 2020 to June 2021, he was a postdoctoral researcher with the Electrical Engineering Department at King Saud University. From 2008 to 2011, he was an electrical engineer at Yemen Mobile Company, Yemen. Since 2011, he has been a lecturer (on leave) with the Electrical Engineering Department at Sana’a University. His research interests include AI-based control of grid-connected renewable energy power plants, microgrids, energy management, power system operation and control, and power system protection.