

Experimental analysis for diagnosis and fault detection in the Internet of Things: A home refrigerator case study

César M. Zapata, Gilberto Osorio-Gómez, and Ricardo Mejía-Gutiérrez
 Design Engineering Research Group (GRID)
 Universidad EAFIT
 Medellín, Colombia
 {czapat34, gosoriog, rmejiag} @eafit.edu.co

Abstract—Implementation of modern data analysis techniques can result in novel applications for the Internet of Things (IoT) far beyond of connectivity of objects. Machine learning could be considered as a mean of generating value in IoT products and this will be shown through a case study of a home refrigerator with a remote Fault Detection and Diagnosis (FDD) mechanism. This document presents the experimental analysis and the results of applying supervised learning and clustering for identification of faulty modes. Finally, the impact of the amount of used sensors on classification accuracy has been analyzed.

Index Terms—Internet of things (IoT), Machine learning, Refrigerator.

I. INTRODUCTION

THE Internet of Things (IoT) is a concept with many years of development that is rapidly gaining ground in the scenario of modern wireless telecommunications, considering the pervasive presence of a variety of things or objects with the possibility of data acquisition, connection and action, such as Radio-Frequency Identification (RFID), sensors, actuators, mobile phones, etc. in day to day life [1]. Furthermore, it is not just about connectivity, but about how such connectivity can be used to improve the behaviour of products, and to learn from their context. Thus, the IoT relies on data analysis to exploit its true potential, and one of its main objectives is to enhance products by exploiting all gathered data [2]. In this way, the potential of any IoT product may not be achieved if possibilities and limitations of data analysis are not considered.

Although several IoT platforms have data analysis libraries and frameworks (ThingSpeak, ThingWorx, AWS IoT, IBM Watson IoT and Google IoT Solutions), data analytics applied to IoT products is in its infancy (i.e. applications usually do not go beyond remote monitoring, descriptive statistics and simple alerts or notifications). Furthermore, many IoT applications still rely on humans for a cognition process [3].

A careful examination of modern data analysis techniques (i.e. expert systems, machine learning, etc.) must be performed in order to make such IoT developments successful [3], [4]. Among such techniques, machine learning offers the processing capabilities demanded by sophisticated IoT applications due to its scalability and adaptability.

This article focuses on machine learning as a mean of exploiting opportunities of sophisticated IoT applications, specifically on a remote Fault Detection and Diagnosis (FDD) system for a standard home refrigerator. Home refrigerators connected to the IoT represent multiple sources of data, such as energy consumption and temperature signals. Besides, considering that this type of appliances are not generally well maintained [5], an important contribution, and product enhancement, is achieved by developing a remote FDD system in order to reduce maintenance costs and to extend equipment lifespan [6].

Since several root causes could have the same effect on the performance of home refrigerators, current maintenance practises demand a considerable amount of time. Technicians rely on custom protocols based on previous experiences and experts to find the root cause. In the past, such practises were the only viable option; but, the development of the IoT offers an incredible opportunity to enhance daily life products and reduce labour costs [7]. In this way, a remote FDD system could reduce the time spent to identify faults and may create opportunities to optimize operation of home refrigerators. Besides, there is not a reported complete work aiming to create a remote FDD system for home refrigerators. This paper is devoted to provide insights to such development based on an analysis of the application of machine learning techniques in the IoT.

Section II presents work related to the application of machine learning to the IoT, and also work related to FDD systems for home refrigerators. Section III presents the case study provided by a local manufacturer of home appliances. This section also presents the first approach to have a remote fault detection. Section IV describes the experimental setup (i.e. the sensor layout) oriented to the development of the FDD system. Section V shows the results and analysis of applying supervised and unsupervised (k-means) learning techniques for the case study. Finally, conclusions and future work are presented in Section VI.

II. RELATED WORK

The first part of this section presents examples of IoT applications developed using machine learning techniques,

whereas the second part is concerned with Fault Detection and Diagnosis (FDD) for home refrigerators.

A. Machine Learning for the Internet of Things (IoT)

Machine learning systems produce output based purely on data (unsupervised learning), or, on data and output examples (supervised learning), in contrast to more traditional computer programs which produce an output based on a set of instructions. Regardless of the approach, any technique could be used as a final application or as a component of a broader development.

Unsupervised learning could be used to discover novelty aspects embedded in the product context, and it is usually used when there is no previous information of a particular aspect in a product. All the same, there are a number of ways to exploit the developments of unsupervised learning in the IoT:

- **Clustering:** this technique is used when the goal is to find different ways in which different sources of data are similar (or dissimilar). It can be used to identify different types of users; uncommon behaviour among users; state of machinery, etc. It has been applied directly to recognize human activities [8], [9]; and indirectly used to schedule home appliances based on their similarity [10].

A technique of k-means clustering involves dividing a set of data points into non-overlapping groups, or clusters, of points, where points in a cluster are more similar to one another than to points in other clusters. When a dataset is clustered, every point is assigned, through an iterative process, to some cluster, and every cluster can be characterized by a single reference point or cluster centroid defined as the mean value of the dataset within the cluster [11], [12].

- **Collaborative filtering:** this technique is used when certain attributes of a user should be predicted based on similarities with other users. For instance, users of similar cars and driving schedules may be interested in the same products. Similarly, Lee et al. [13] have used it for user group recommendations. Kim et al. [14] developed another application using collaborative filtering to estimate depression index in a smart health scenario.

As opposed to unsupervised algorithms, supervised learning could be used only when it is clear the expected output; in fact, it is required to provide some examples of what is expected. Its usages for the IoT can be broadly classified as follows:

- **Regression:** this technique is used when the goal is to predict a real value. Garcia et al. [15] used it to predict travel time for buses; and the power consumption of electric vehicles trips was predicted by Zheng et al.[16].
- **Classification:** this technique is used when the goal is to distinguish between groups of data patterns. A given situation can be seen as a classification problem when there is a finite number of possible values that can be assigned to the data being analyzed. Classification techniques have been used by Madeira et al. [17] to detect presence of humans without using cameras, and by Wu et

al. [18] to classify human activity in the context of smart homes.

There are certain scenarios where an unsupervised learning technique can be used for the same purpose of a supervised one, though for those cases, supervised learning can be more accurate. Moreover, an application of any machine learning approach can be applied simultaneously with others using the same or different data as long as there are resources to compute. In fact, Yuan et al. [19] used a mixed approach called semi-supervised learning, to estimate water quality levels.

B. Fault Detection and Diagnosis (FDD) for Home Refrigerators

Real-time FDD has been an area of interest in several industries [20], [21], [22], [23], and has been applied for the last two decades [24] for Heating, Ventilating, Air Conditioning (HVAC) systems [25], [26], [27]. However, its application only relates to industrial and commercial scenarios.

Authors have found that very little work has been dedicated to FDD for food handling systems, including refrigerated containers, supermarket refrigerators and home refrigerators. This could be explained by the relative high cost of sensors and computing power. Moreover, most research on FDD has been conducted in isolated subsystems and components within the refrigeration system (e.g. condenser unit, compressor, etc.). Although one can argue that the refrigeration cycle in food handling systems and other HVAC is very similar, operation occurs over a different range of temperatures and the systems utilize different refrigerants [5]. The most relevant works related to the case of FDD of home refrigerators are presented here.

One approach that has been considered is the FDD based on the system model. Nevertheless, system modelling becomes more complex in relation with the complexity of the system [28], and it has been performed only on subsystems, such as the condenser [28] or the compressor [29]. Yang et al. [30], [31] focused on the evaporator part of the refrigeration system; their model allowed to distinguish between faults at the component (evaporator) and sensor level. Regardless of the sensor readings used to account for the faults considered, these approaches rely on empirical fine tuning and estimation of thresholds and other parameters present in system's equations. These may vary greatly when the refrigeration unit is placed in other geographical context. Therefore, such approach can not be effectively applied in the context of IoT.

A more suitable approach in the context of the IoT is known as data-driven approach. Two different works [32], [33] used energy consumption measurements for fault detection. However, Mavromatidis et al. [32] were concerned with main energy consumers in a supermarket, so the goal in that project was not to isolate individual faults in the food handling system, but their procedure identified the subsystem that, due to faulty operation, was causing a deviation from an energy consumption model created by artificial neural networks. On the other hand, Srinivasan et al. [33] were focused only in the refrigeration system, and more detail was extracted from the energy consumption readings. They isolated the faults based

on the way in which energy signals deviated from a base line built by using an statistical model.

Wichman et al. [5] used several sensors supported by a system description on the physical level, in a mathematical form, to provide additional descriptors of the system and faults. Although their approach still relies on threshold values, it is important to highlight that some known physical relationships can help to eliminate the need of additional sensors on the system; reducing costs in the final massive application. Their work was performed on a restaurant's walk-in cooler, which was more related to home refrigerators than previous work. Nevertheless, the dynamics of this kind of products are clearly different from those most commonly found in homes.

Authors only found one research, proposed by Horng et al., tackling the problem of FDD in home refrigerators [34], but such research was focused on the development of a web platform in which the fault detection mechanism was used. They did not detail the faults considered nor the effectiveness of their approach, although their work is relevant in the context of the IoT. Their approach was based on a finite state machine algorithm, which relies on hand-coded knowledge of threshold values for every signal. This approach is very limited, since perfect knowledge of the system and its interaction with the environment is often unavailable. Moreover, home refrigerators are used globally, implying that their environment is not always the same or even similar.

To summarize, there is not a reliable FDD system built for home refrigerators, so the first step in such development needs to be experimental. Food handling systems have economical limitations regarding the number of affordable sensors to be placed in the system; therefore the problem of FDD should be more carefully considered.

III. CASE STUDY: HOME REFRIGERATORS

Home appliances manufacturers generally provide after-sales service, such as warranty of their products, which often goes from one year up to ten years. Warranty services make manufacturers to incur in considerable costs for massively produced products, such as home refrigerators, specially considering that domiciliary technical services are usually required.

In order to have a home refrigerator connected to the IoT, a remote FDD system could be built to significantly reduce maintenance costs by real-time fault diagnosis, helping technicians to work more efficiently. Moreover, any faulty condition could even be detected before users report it, and optimal operation condition would be assured, concurrently reducing energy consumption rates, and extending equipment lifespan [6].

The diagnosis of home refrigerators is very time consuming, since it relies on checking for malfunction of every component in the product, following a defined protocol. Furthermore, several faulty conditions may reflect the same consequences for users. Most common faults reported by the local manufacturer are:

- **Refrigerant Leakage.** As a result of this fault the refrigerator does not perform accordingly to its design. The physical phenomenon determined by the user is abnormal

warm temperatures at the freezer and fresh food compartments. When this fault is not reported on time, in most of the cases, storage food inside the refrigerator is no longer available for human consumption due temperature rising for long period of time. A possible cause of this issue is related to abnormal welding procedure on each pipe junction along the refrigeration system, for instance the pipe junction between the compressor suction and warm frame pipes, between warm frame and condenser pipes, between condenser discharge and heat exchange pipes, and so on. Also, this fault could be related to an improper refrigerator handling procedure during transportation.

- **Compressor Failure.** This issue causes internal temperature rising of both fresh food and freezer compartments inside the refrigerator due to malfunction of the compressor. In most of the cases, this fault is related to compressor coils burn out, or coils being short circuited. Malfunction of the coils is difficult to predict since this issue is usually related to the lifespan of the compressor. Another important reason of compressor malfunctions is caused by its internal valves. Internal compressor valves got defective by the action of continuous work of the compressor, and valves are affected by mechanical wear of the refrigerant fluid compression process.
- **Fan Damage.** Air fan motor issues cause slightly increase of the temperature in both compartments. Nonetheless, it is more critical at the fresh food compartment due to not cold air circulation by the effect of the movement of the air fan blades. Most of the cases are related to burn out of the coils, coils been short circuited, or locked fan blades.
- **Pipe Clogging.** This issue could occur because of the impurities or residues during the manufacturing process, due to an inadequate vacuum process unable to remove humidity circulating along the refrigeration pipes creating an ice clog at the capillary and non-condensate gases of the refrigeration system [35]. Impurities could remain also after a non-proper repair procedure after sales, but, abnormal repair procedures are rare to happen since these are executed by qualified personnel. Other cause is not proper handling procedures causing damages to the refrigeration pipe's system, where the worst case scenario is the restriction of the refrigerant fluid to the capillary pipe. Dryer filter also can be clogged by the effect of the compressor oil, in case the product is being lay-down toward the compressor line service causing compressors oil filling the refrigeration pipe's system. Also, the silica (component inside dryer filter) can clog the capillary pipe.
- **Loose Door.** This issue could be caused during the manufacturing process, transportation or product handling. Loose door allows warm air to enter the fresh food or freezer compartments, causing temperature rising in these compartments.
- **Broken Dampers.** This issue represents a non-control of the cold air from the freezer compartment to the fresh food compartment by the effect of broken damper. This issue could be reported by the user due to either cold temperatures or warm temperatures at the fresh food compartment. Strong manipulation of the knob that is

assembled with the damper can easily break this component. This issue can be associated with the conditions of the defrost cycle. The melted ice converted in water can stuck the damper, so when the user tries to move the damper, it can be deformed, bent, or broken easily.

Table I summarizes these faults, their perceived effect by users, and an estimation of the time spent by technicians to detect the faulty condition.

TABLE I: Most common faults.

Faulty condition	Perceived consequences	Time for detection (min)
Refrigerant Leakage	Conservation area is warmer than usual. Freezer does not freeze.	30
Broken damper	Conservation area is colder than usual.	10
Loose doors	Conservation area is warmer than usual. Freezer does not freeze. Internal condensation.	30
Compressor failure	Conservation area is warmer than usual. Freezer does not freeze. Noise.	25
Fan damage	Conservation area is warmer than usual. Freezer does not freeze.	20
Pipe clogging	Conservation area is warmer than usual. Freezer does not freeze.	30

Faults presented in table I, and others, are related to each of the components of the home refrigerator shown in figure 2. However, the development of some faults can affect multiple components at the same time.

The local industrial partner provided a report with the number of warranty services that were required for home refrigerators in 2015 (see figure 3). A total of 49,786 domiciliary services across the country were reported, costing over USD \$ 2 million. Although this value corresponds to a small fraction of all sales, each reported service required technicians to diagnose the product domiciliary. Moreover, some of these requested services could have been solved by providing relatively simple instructions to the user about the operation of the product.

It is also important to note that the scale and the market of home refrigerators impose several constrains that are not present in other contexts. For example, adding a considerable amount of sensors to the product may increase the list price of the product, making it non-competitive for current users. So, it is of great interest to provide great accuracy in the FDD system while using a minimal amount of sensors.

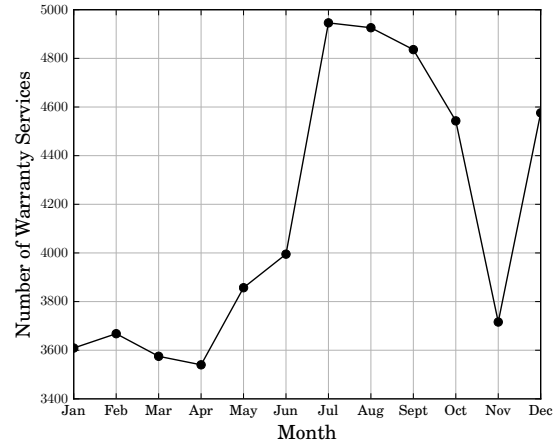


Fig. 3: Number of warranty services during year 2015.

At the last part of the case of study is presented a configuration of smart refrigerator and the initial capabilities to communicate a malfunction of its actuators as fault detection. The implementation of the machine learning algorithms either in the firmware of the refrigerator embedded systems or in remote server for data acquisition is still under development and it will be considered as future work.(see figure 1) The experiment provides an engineering development from identifying requirements, design, prototype, and test to commercialisation and operations. The aim of using IoT in a smart home is to optimise electricity usage, offer personalisation, enabling advanced maintenance, and affordable system upgrades with higher customer satisfaction. This case study was dedicated to mass production of ordinary refrigerators in Medellin, Antioquia. This effort is part of a portfolio, including several IoT-enabled home appliances projects in an established factory with over 70 years of operations. Converting ordinary home appliances to smart ones is highly demanded concerning the minimum impact on current production lines. Designing a board for a conventional appliance requires seamless integration between sensors, actuators and electrical systems. This refrigerator was manufactured to use a central control board. Therefore, a new IoT board has been designed for working between the refrigerator,user and potentially the factory. The current system board has been kept without any change to save the cost at the lowest possible. The new PCB (printed circuit board) for IoT has to be placed into a small space while preserving reliability and product budget. Flexible PCBs have emerged as the best strategy for IoT-enabled R&D projects. With support for flexible PCB design, comprehensive testing tools, and an inbuilt interface, Altium as a reliable, trustworthy software has been used. The component placement was optimised for space and performance based on the mechanical constraints on the top right side of the refrigerator's door. The PCB shape was modified to accommodate the board in the place. After the validation of the correct functionalities,several improvements have been considered including 1) better connectors for AC signals 2) DC and AC decoupling 3) Wireless (Wi-Fi and Bluetooth) antenna position.

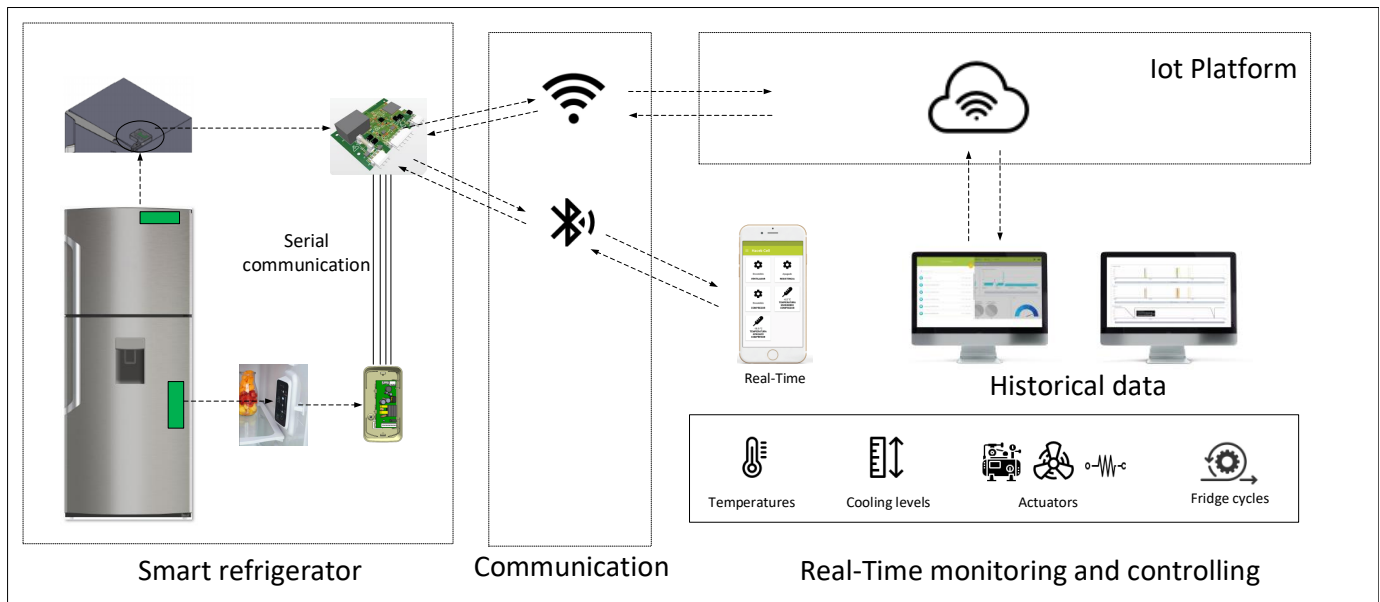


Fig. 1: End-to-end IoT-enabled smart home appliance system.

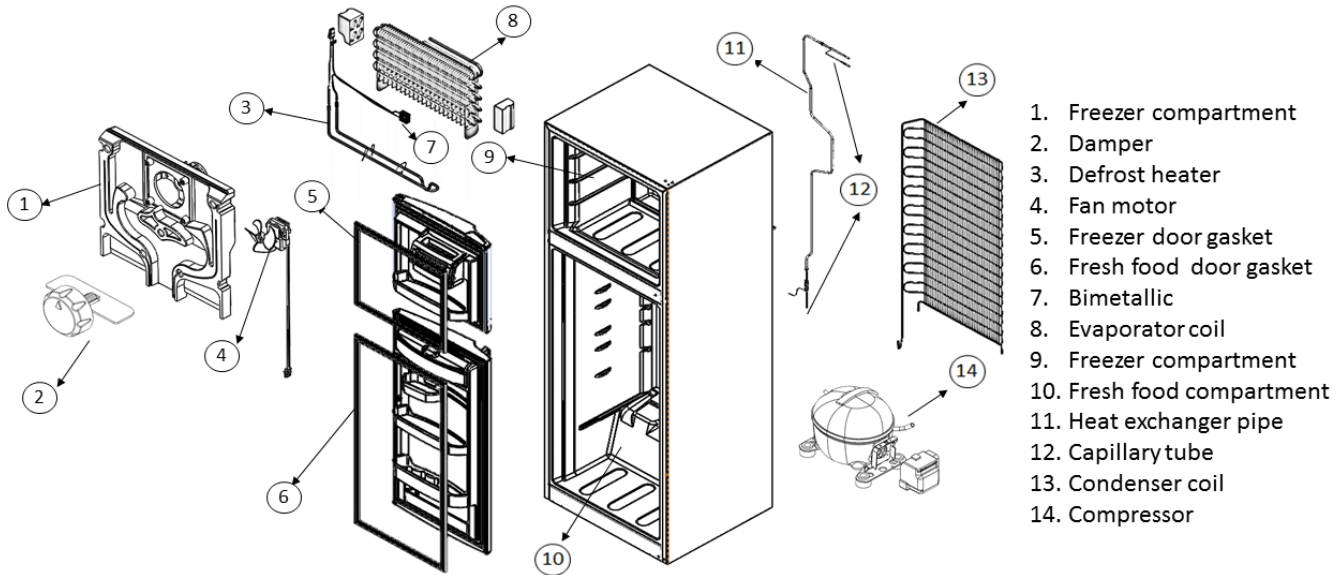


Fig. 2: Description of a refrigerator produced by local partner since 2012.

IV. EXPERIMENTAL SETUP

In order to identify the signals of both a functional and a broken refrigerator and to determine the required number of sensors, a data acquisition system was implemented into a refrigerator with a total number of 17 temperature sensors. Its layout is shown in Figure 4 and Figure 5, and detailed in table II.

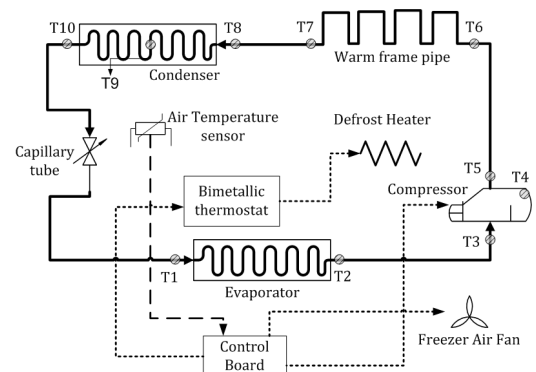


Fig. 4: Location of sensors in the system.

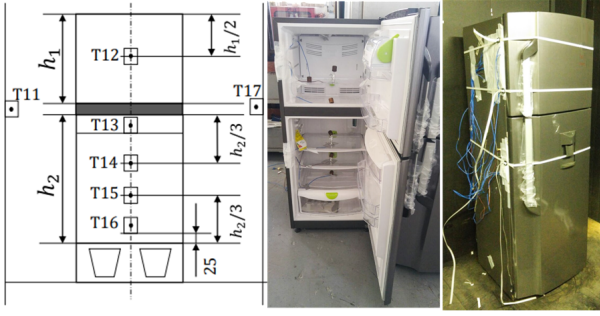


Fig. 5: Location of sensors inside the product compartments. h_1 and h_2 refer to the height of the freeze and conservation areas respectively.

TABLE II: Sensor labels and location.

Sensor label	Location
T_1	Evaporator coil input
T_2	Evaporator coil output
T_3	Compressor suction input
T_4	Compressor frame
T_5	Compressor discharge output
T_6	Warmer frame pipe input
T_7	Warmer frame pipe output
T_8	Condenser coil input
T_9	Condenser medium section
T_{10}	Condenser coil output
T_{11}	Room temperature 1 (right side of the product)
T_{12}	Medium section of the freezer compartment
T_{13}	Quick chill compartment
T_{14}	Top section of fresh food compartment
T_{15}	Medium section of fresh food compartment
T_{16}	Bottom section of fresh food compartment
T_{17}	Room temperature 2 (left side of the product)

LabVIEW (National Instruments, Austin, TX) is a data acquisition and programming environment that allows flexible acquisition and processing of analog and digital data. LabVIEW was implemented in the temperature controller refrigerator test chamber as a tool to data acquisition of the temperatures to determine the proper performance of the refrigerator according to the standard IEC62552.

From the point of view of machine learning, the Fault Detection and Diagnosis (FDD) task could be seen as a supervised learning task, in which there is (or can be built) data associated with each faulty condition.

This research considers only three of the main failures presented in the previous section: refrigerant leakage, broken damper, and loose doors. Each fault was recreated, and data was recorded for over 24 hours.

Experiments for normal operation (i.e. recorded data of the refrigerator working properly) were also carried out. So, in terms of supervised learning, there were four different classes or labels. Before the experiments, the refrigerator was checked for proper work according to the manufacturer standard.

Although home refrigerators could be exposed to a great variety of surrounding temperatures (i.e. they can be placed in different cities and should perform well regardless of the weather), all the experiments, except no load pull down test, were carried out in a room with controlled temperature and

humidity, at 32°C and 55% respectively, according with the IEC 62552 standard.

V. RESULTS AND ANALYSIS

After implementation of sensors in the refrigerator, it was necessary to validate the proper assembly conditions of the refrigerator. No-load pull down test was performed to validate correct temperature conditions along the refrigeration system. The test was executed, for 24 hours, at ambient temperature of 43°C and relative humidity of 75% \pm 5% [36]. The temperature controller of the product was taken apart to allow the product running without compressor cycling by the effect of determined sensor threshold temperature values to turn either the compressor ON or OFF. The cooling speed is measured by the time it takes until the average temperature of fresh food compartment reaches 5°C and freezer compartment -18°C with no load inside and with all the doors closed.

When the compressor was turned on the temperatures T_1 , T_2 , T_{12} , T_{14} , T_{15} , T_{16} started to decrease. It is worth noting that temperatures for T_1 and T_2 should be almost identical after a couple of hours of the test as this is a design condition. T_9 needs to be either 5°C above or below room temperature test conditions. The results of this test showed a refrigerator in good operation conditions according to its design, as it is presented in Figure 6.

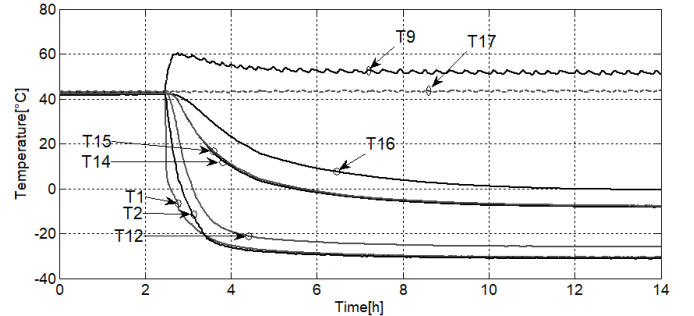


Fig. 6: No-load pull down test. Temperature vs. Time

Firstly, results of applying supervised learning algorithms for FDD in home refrigerators are presented. All sensors were considered for the analysis; however, using the whole set of sensors will dramatically increase the product cost. Therefore, additional measures of the effectiveness of the algorithms were taken in sub sets of the sensor readings.

Removing outliers, filtering and balancing the whole data set (i.e. the combination of all experiments' data) each experiment resulted in 1661 observations of 17 sensor readings each. The whole data set to form both the training and test sets was randomly split.

Table III shows the performance of several classifiers using all sensor readings. The training accuracy is reported using the mean of 10-fold cross validation accuracy, whereas in the test set it is simply the overall accuracy. Furthermore, every classifier was trained using grid search to find the corresponding set of best hyper-parameters. Great accuracy is observed in all of them.

Nonetheless, as explained before, it is not feasible to deploy such a big number of sensors in a real application. Moreover,

TABLE III: Classifier's performance on full data set.

Method	Overall accuracy (%)	
	Training	Testing
K-Nearest Neighbours	99.25	95.79
Logistic Regression	100	96.21
Neural Networks	99.70	97.26
Support Vector Machines	100	95.64

the experimental set up was redundant in the sense that many variables are highly related, so it should be possible to learn patterns of faulty modes from a reduced number of sensors.

Ten data sets with k number of elements were created, where each data set had a number of variables chosen from the combination $\binom{17}{k}$ and $k = \{1, 2, 3, \dots, 15\}$. Logistic Regression and Neural Networks were trained and tested using each data set for each number of sensors, and the mean and standard deviation of their accuracy were recorded on the training and test sets.

Figures 7 and 8 show the mean and standard deviation of the results using the training and test sets respectively. Only the results of the best two algorithms from table III, Neural Network and Support Vector Machine, were considered.

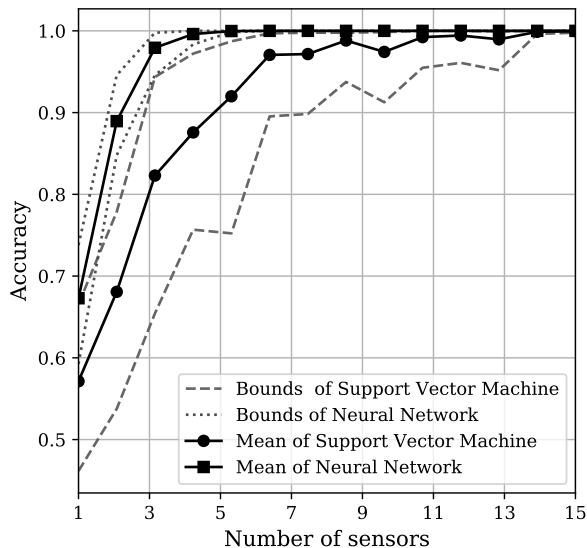


Fig. 7: Performance of two classifiers in the training with different number of sensor.

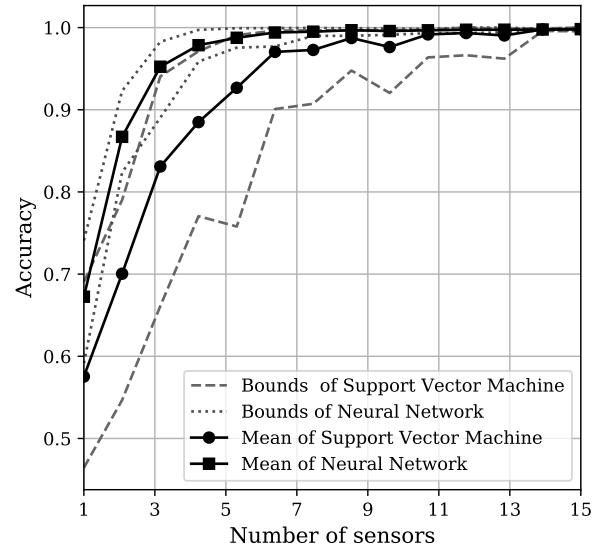


Fig. 8: Performance of two classifiers in the test with different number of sensor.

It is observed that the number of sensors affects considerably the accuracy of the both Logistic Regression and Neural Networks methods. However, high accuracy was achieved by both methods with as little as seven sensors.

A further test for the classifiers performance, using a data set built by a combination of two days of normal operation, was carried out. The difference between this new test set and previous one is that the first one was built with the same experiments that produced the training set, so it is expected to perform better on that set. On the other hand, the training set did not have any data generated by the experiments in the second test set. Figure 9 shows that the performance of both methods is worse with respect to the other data sets, indicating that although a viable solution can be developed, a more comprehensive data set should be built including information from repeated experiments.

After the implementation of supervised machine learning techniques, k-means clustering, as a unsupervised machine learning algorithm, was proposed. The aim was the analysis of the proper performance of the product (healthy) vs. the performance of the product having a malfunction (faulty) employing less amount of sensors. In order to validate that the product is operating correctly, the refrigerator was exposed, for 45 hours, to a room temperature condition of 32°C . In this experiment, only the temperatures $T_1, T_2, T_9, T_{12}, T_{14}, T_{15}, T_{16}$, and T_{17} were considered. In this way, Figure 10 represents the normal performance of the refrigerator.

In order to describe the performance of the refrigerator exposed to a room temperature condition of 32°C having the failure related to gas leakage, the refrigerator was charged with 30g of R-600a refrigerant fluid, equal to 27% less of its design R-600a charge (41g). This was the way to simulate the fault, due to the technical difficulties to replicate a fault by the effect of a micro-leakage. Data was recorded about 70 hours, however, for the cluster comparison it was taken the same data recording time, near to 45 hours. As well, in this

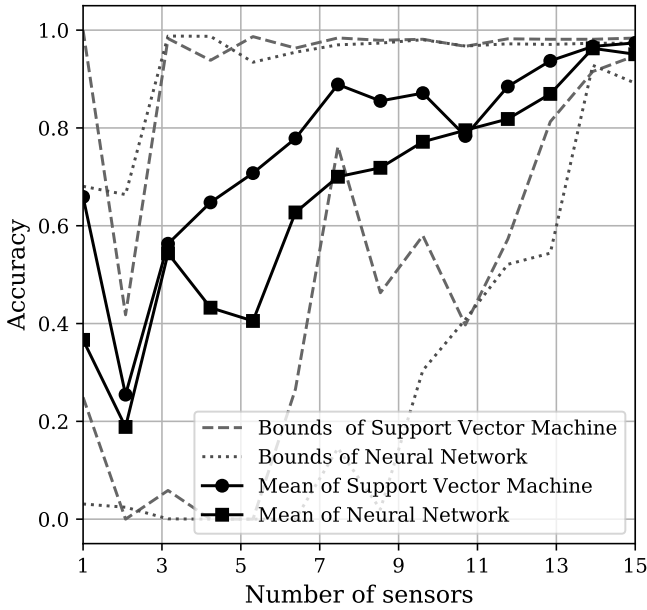


Fig. 9: Performance of two classifiers in a second test set with different number of sensors.

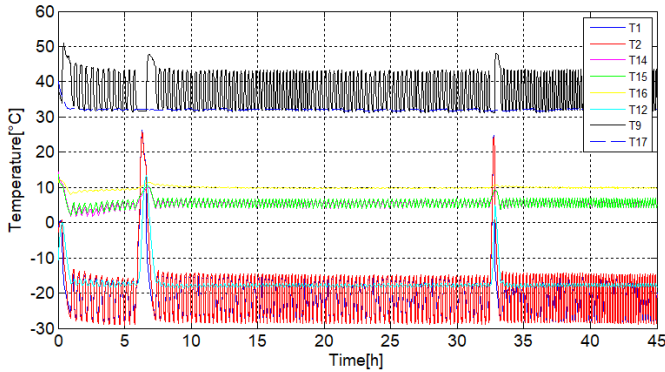


Fig. 10: Refrigerator performance without failure. Temperature vs. Time

experiment only the temperatures T_1 , T_2 , T_9 , T_{12} , T_{14} , T_{15} , T_{16} , T_{17} were considered. In this way, Figure 11 represents the temperature of different spaces into the refrigerator when there is a gas leakage.

The information of figures 10 and 11 was treated in order to quantize the information in a data set of RMS values to apply the analysis of the best centroid values calculated by a k-mean algorithm. Quantization process was applied to temperatures data set of T_2 , T_9 , and T_{15} and it consisted in classifying the dependent variable in 200 intervals in order to treat the information. The independent variable was processed by reducing five consecutive values of the continuous data set to a discrete single value. Figure 12 represents the overlapped quantized temperatures of healthy and faulty cases. It seems that the temperatures data set for healthy product has more recorded data, however the recorded time for both tests was the same. The difference observed in the figure 12 is related to record data frequency of the acquisition data system used. Comparing

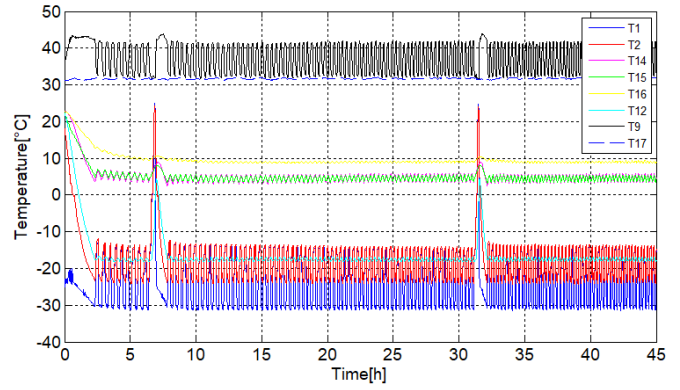


Fig. 11: Refrigerator performance with failure related to less quantity of refrigerant fluid in the refrigeration circuit. Temperature vs. Time.

both results, it is shown that refrigerator performances are almost identical and it is difficult to determine either the product has a malfunction or not. In this way, to determine a FDD becomes relevant. In addition, the data set treatment was considered as a feasible alternative to be implemented in mass production due to only three sensors would be necessary to be included in the bill of materials for the evaluated refrigerator model.

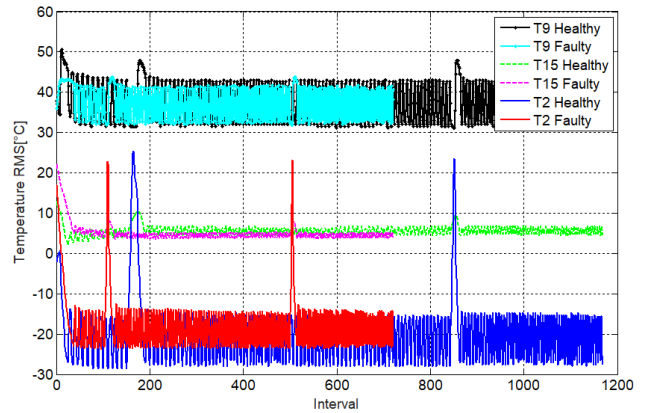


Fig. 12: RMS temperatures: healthy vs faulty

For the analysis, 100 clusters were used to acquire the centroids for both healthy and faulty products. Results were also compared to determine the differences between both sets of centroids found. It is observed, in Figure 13, that centroids distribution of healthy vs faulty are completely different.

The use of k-means strategies is relevant for this case study since it allows to classify two data sets that were almost identical in two different data sets that can be analyzed by subsequent tools. K-means also helps to filter the information and to improve the visualization of the data set. It is remarkable that the use of k-means strategies request less amount of data.

K-means offered an excellent method for filtering a whole data set, however, the results were processed to determine the level of accuracy of the treated data set. The differences observed in the centroids allowed the treatment of the infor-

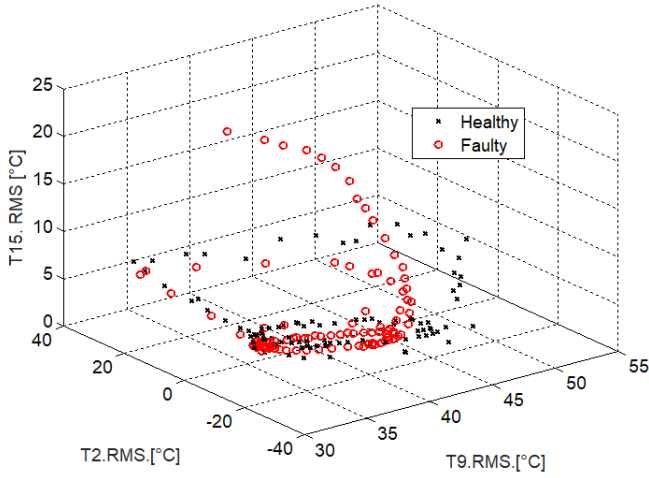


Fig. 13: Centroids comparison. RMS data set.

mation as a supervised classification method since it uses the class labels of the training data. K-nearest neighbours (KNN) algorithm was used for the analysis of the K-means RMS treated data sets.

Table IV shows the performance of KNN classifier using the information obtained from K-means treated data set. The training was executed using K values from one to ten ($K = \{1, 2, 3, \dots, 10\}$). During the testing of the classifier, its accuracy to identify a healthy or faulty performance is pretty high, specifically for the data set obtained from the three sensors (T_2 , T_9 , and T_{15}) that were analyzed.

TABLE IV: Performance of the classifier based on K-means data set.

K tested value	Overall accuracy of KNN
	Testing(%)
1	95.00
2	90.00
3	95.00
4	90.00
5	82.50
6	90.00
7	85.00
8	90.00
9	82.50
10	82.50

- The real-time monitoring and notification** Smart IoT monitoring from the edge to the Cloud provide unique user experiences by seeing real-time information and advanced analytics result on a dashboard. The smart refrigerator has enabled to leverage AI-driven big data analytics for automated fault detection, root cause analysis and personalisation. Providing visibility into a smart home across all IoT-enabled appliances, analyse user data in real-time to measure satisfaction, detect struggles, predict and offer personalized options. Real-time monitoring capabilities, including Cloud and IoT. Ubidots is an IoT development platform that allows users to visualise and analyse the data received from connected devices. Ubidots has different tools to extract valuable information

from the collected data. The smart refrigerator dashboard has line charts useful to visualize both digital and analogue time-based variables. Fig. 14 visualised fresh food and freezer temperatures by using line charts and the instant state of the compressor, defrost heater, and air fan. Also, the cooling level visualised by using a gauge indicator.

VI. CONCLUSION AND FURTHER WORK

Connectivity is not the goal of the IoT, the goal is the use of such connectivity for the development of novel applications. Such development could only be achieved by analysis of data generated from interconnected objects. Therefore, an examination of data analysis techniques must be considered during the design and development of novel applications.

In this article the importance of the experimental phase in the development of a novel IoT application is highlighted. This development went through the steps of finding an opportunity and testing solutions. The faults considered in the previous section do not cover the entire set of faults reported by the industry partner, and further experiments must be carried out. Nevertheless, obtained results suggest that a remote FDD system for home refrigerators could be built using machine learning with a certain number of sensors, and this would take to reduce times of fault detection, reported in Table I, and to improve efficiency of maintenance. The location and exact amount of such sensors, which may greatly vary, could only be derived when having a data set consisting of all possible faults, but using the procedure here described could be very useful.

It was noticed that, although the algorithms performed well on data from the same day, generalization to data sets from different days is harder with a limited number of sensors, but the use of more sensors shows the potential for generalization across data from different days. Further work will include a more comprehensive data set, which should be built using several refrigerators with different room temperatures; several days worth of data; the generation of additional variables for each experiment based on physical laws; and an analysis of the feasible networking alternatives for a final deployment.

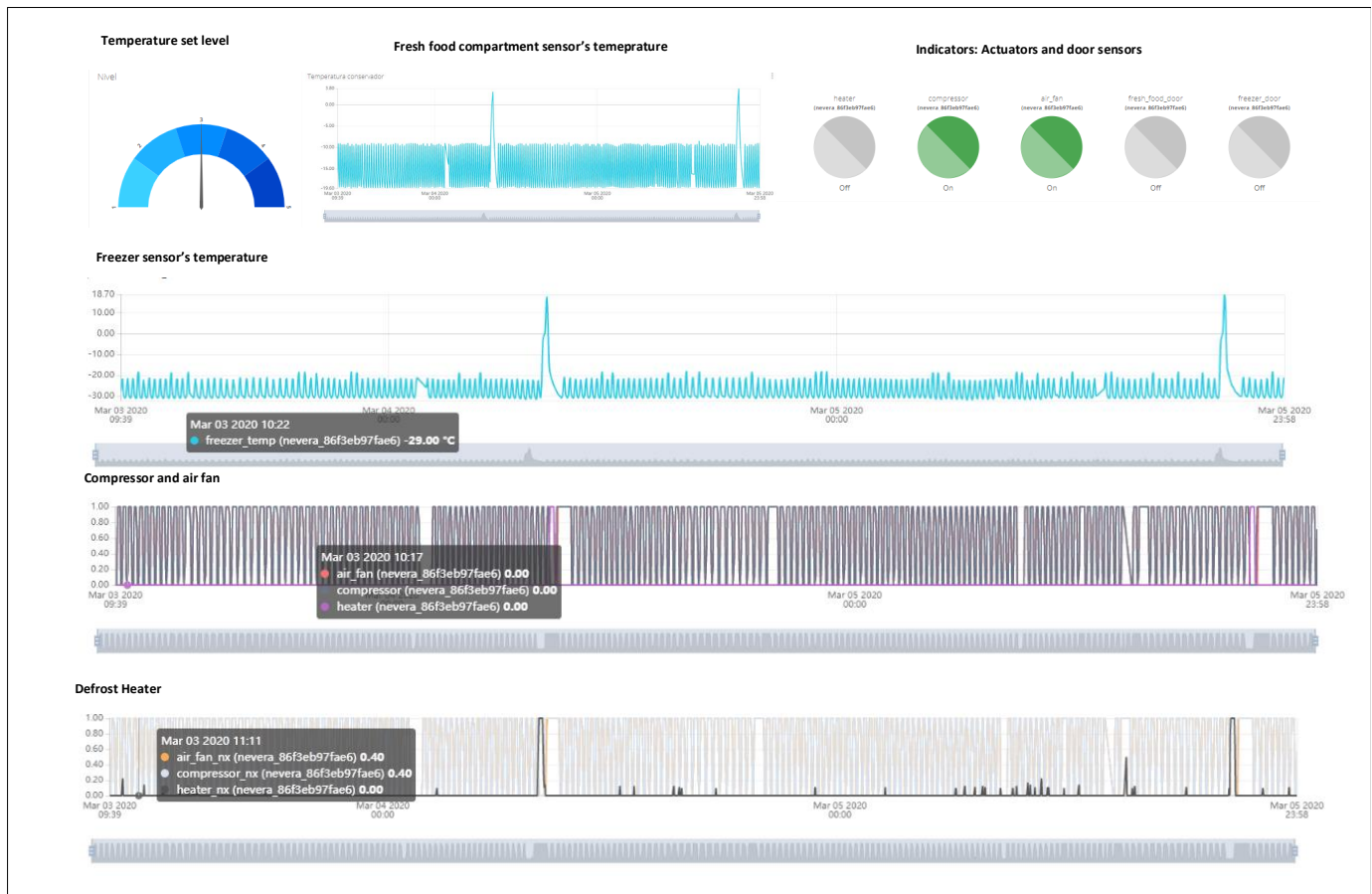


Fig. 14: Fresh food and freezer temperatures with state of actuators including compressor, air fan, defrost heater, fresh food and freezer doors..

REFERENCES

- [1] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1389128610001568>
- [2] O. Vermesan, P. Friess, P. Guillemin, H. Sundmaeker, M. Eisenhauer, K. Moessner, F. Le Gall, and P. Cousin, *The Internet of things*. River Publishers, 2013.
- [3] Q. Wu, G. Ding, Y. Xu, S. Feng, Z. Du, J. Wang, and K. Long, "Cognitive Internet of Things: A New Paradigm Beyond Connection," *IEEE Internet of Things Journal*, vol. 1, no. 2, pp. 129–143, 2014. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6766209>
- [4] C. W. Tsai, C. F. Lai, M. C. Chiang, and L. T. Yang, "Data mining for internet of things: A survey," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 1, pp. 77–97, 2014.
- [5] A. Wichman and J. Braun, "Fault Detection and Diagnostics for Commercial Coolers and Freezers," *HVAC&R Research*, vol. 15, no. 2007, pp. 77–99, 2009.
- [6] P. Sreedharan and P. Haves, "Comparison of Chiller Models for Use in Model-Based Fault Detection," *International Conference for Enhanced Building Operations*, no. June, 2001.
- [7] Y. Huang and J. Zhan, "GA and RBF based real-time FDD for refrigeration units," in *2009 International Symposium on Intelligent Ubiquitous Computing and Education, IUCE 2009*, pp. 22–25, 2009.
- [8] J. Ye, L. Fang, and S. Dobson, "Discovery and recognition of unknown activities," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, 2016, pp. 783–792.
- [9] S. A. Fahad and M. M. Alam, "A modified k-means algorithm for big data clustering," *International Journal of Science, Engineering and Computer Technology*, vol. 6, no. 4, p. 129, 2016.
- [10] M. Rasheed, N. Javaid, M. Awais, Z. Khan, U. Qasim, N. Alrajeh, Z. Iqbal, and Q. Javaid, "Real time information based energy management using customer preferences and dynamic pricing in smart homes," *Energies*, vol. 9, no. 7, p. 542, 2016.
- [11] V. Faber, "Clustering and the continuous k-means algorithm," *Los Alamos Science*, vol. 22, no. 138144.21, p. 67, 1994.
- [12] X. Luo, K. Fong, Y. Sun, and M. Leung, "Development of clustering-based sensor fault detection and diagnosis strategy for chilled water system," *Energy and Buildings*, vol. 186, pp. 17–36, 2019.
- [13] O.-J. Lee, M.-S. Hong, J. J. Jung, J. Shin, and P. Kim, "Adaptive collaborative filtering based on scalable clustering for big recommender systems," *Acta Polytechnica Hungarica*, vol. 13, no. 2, pp. 179–194, 2016.
- [14] J. Kim, V. D. Calhoun, E. Shim, and J.-H. Lee, "Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia," *Neuroimage*, vol. 124, pp. 127–146, 2016.
- [15] F. C. C. Garcia and A. E. Retamar, "Towards building a bus travel time prediction model for metro manila," in *2016 IEEE Region 10 Conference (TENCON)*. IEEE, 2016, pp. 3805–3808.
- [16] B. Zheng, P. He, L. Zhao, and H. Li, "A hybrid machine learning model for range estimation of electric vehicles," in *2016 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2016, pp. 1–6.
- [17] R. Madeira and L. Nunes, "A machine learning approach for indirect human presence detection using iot devices," in *2016 Eleventh International Conference on Digital Information Management (ICDIM)*. IEEE, 2016, pp. 145–150.
- [18] C.-L. Wu, Y.-S. Tseng, and L.-C. Fu, "Spatio-temporal feature enhanced semi-supervised adaptation for activity recognition in iot-based context-aware smart homes," in *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*. IEEE, 2013, pp. 460–467.

- [19] D. Yuan, J. Richardson, R. Doherty, C. Evans, and E. Altendorf, "Semi-supervised word sense disambiguation with neural models," *arXiv preprint arXiv:1603.07012*, 2016.
- [20] J. Mohammadpour, M. Franchek, and K. Grigoriadis, "A survey on diagnostics methods for automotive engines," *Proceedings of the 2011 American Control Conference*, vol. 13, no. December 2010, pp. 985–990, 2011.
- [21] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors - A review," *IEEE Transactions on Energy Conversion*, vol. 20, no. 4, pp. 719–729, 2005.
- [22] J. Marzat, H. Piet-Lahanier, F. Damongeot, and E. Walter, "Model-based fault diagnosis for aerospace systems: a survey," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Eng[1] J. Marzat, H. Piet-Lahanier, F. Damongeot, and E. Walter, "Model-based fault diagnosis for aerospace systems: a survey," Proc. Inst. Mech. Eng. Part G J. Aerosp. E*, vol. 226, no. 10, pp. 1329–1360, 2012. [Online]. Available: <http://pig.sagepub.com/lookup/doi/10.1177/0954410011421717>
- [23] H. Henao, G.-A. Capolino, M. Fernandez-Cabanas, F. Filippetti, C. Bruzzese, E. Strangas, R. Pusca, J. Estima, M. Riera-Guasp, and S. Hedayati-Kia, "Trends in Fault Diagnosis for Electrical Machines: A Review of Diagnostic Techniques," *IEEE Industrial Electronics Magazine*, vol. 8, no. 2, pp. 31–42, 2014. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6839118>
- [24] S. Katipamula and M. Brambley, "Review Article: Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems—A Review, Part I," *HVAC&R Research*, vol. 11, no. 1, pp. 3–25, 2005. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/10789669.2005.10391123>
- [25] Y. Yu, D. Woradetchjumroen, and D. Yu, "A review of fault detection and diagnosis methodologies on air-handling units," *Energy and Buildings*, vol. 82, pp. 550–562, 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.enbuild.2014.06.042>
- [26] P. Taylor, T. M. Rossi, and J. E. Braun, "A Statistical , Rule-Based Fault Detection and Diagnostic Method for Vapor Compression Air Conditioners A Statistical , Rule-Based Fault Detection and Diagnostic Method for Vapor Compression Air Conditioners," *HVAC and R Research*, vol. 3, no. 1, pp. 19–37, 2011.
- [27] Y. Zhao, F. Xiao, J. Wen, Y. H. Lu, and S. W. Wang, "A robust pattern recognition-based fault detection and diagnosis (FDD) method for chillers," *Hvac&R Research*, vol. 20, no. 7, pp. 798–809, 2014.
- [28] R. L. Christensen, R. Wisniewski, and K. K. Sorensen, "Safety verification of refrigeration containers using barrier certificates," *Proceedings of the IEEE International Symposium on Computer-Aided Control System Design*, vol. 2016-October, pp. 635–640, 2016.
- [29] D. Dragomir-Daescu, A. A. Al-Khalidy, M. Osama, and G. B. Kliman, "Damage detection in refrigerator compressors using vibration and current signatures," *IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, SDEMPED 2003 - Proceedings*, pp. 355–360, 2003.
- [30] Z. Yang, K. B. Rasmussen, A. T. Kieu, and R. Izadi-Zamanabadi, "Fault detection and isolation for a supermarket refrigeration system—part one: Kalman-filter-based methods," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 13 233–13 238, 2011.
- [31] —, "Fault detection and isolation for a supermarket refrigeration system—part two: Unknown-input-observer method and its extension," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 4238–4243, 2011.
- [32] G. Mavromatidis, S. Acha, and N. Shah, "Diagnostic tools of energy performance for supermarkets using Artificial Neural Network algorithms," *Energy and Buildings*, vol. 62, pp. 304–314, 2013. [Online]. Available: <http://dx.doi.org/10.1016/j.enbuild.2013.03.020>
- [33] S. Srinivasan, A. Vasan, V. Sarangan, and A. Sivasubramaniam, "Bugs in the Freezer: Detecting Faults in Supermarket Refrigeration Systems Using Energy Signals," *Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems*, pp. 101–110, 2015. [Online]. Available: <http://doi.acm.org/10.1145/2768510.2768536>
- [34] M. F. Horng, M. H. Hung, Y. T. Chen, J. S. Pan, and W. Huang, "A new approach based on XMPP and OSGi technology to home automation on web," *2010 International Conference on Computer Information Systems and Industrial Management Applications, CISIM 2010*, pp. 487–490, 2010.
- [35] D. G. Kroger and W. M. Rohsenow, "Condensation heat transfer in the presence of a non-condensable gas," *International Journal of Heat and Mass Transfer*, vol. 11, no. 1, pp. 15–26, 1968.
- [36] C.-C. Yu and T.-P. Teng, "Retrofit assessment of refrigerator using hydrocarbon refrigerants," *Applied thermal engineering*, vol. 66, no. 1-2, pp. 507–518, 2014.